DOCUMENT RESUME

ED 309 196 TM 013 682

AUTHOR	Reisner, Elizabeth R.; And Others
TITLE	Assessment of the Title I Evaluation and Reporting System.
SPONS AGENCY	Department of Education, Washington, DC. Office of Planning, Budget, and Evaluation.
PUB DATE	Apr 82
CONTRACT	300-82-0008; 300-82-0009; 300-82-0010; 300-82-0011; 300-82-0012
NOTE	141p.
PUB TYPE	Collected Works - General (020) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Compensatory Education; Disadvantaged Youth; Elementary Secondary Education; *Evaluation Utilization; Federal State Relationship; *Information Systems; Program Effectiveness; *Program Evaluation; School Districts; School Effectiveness; Self Evaluation (Groups); *Technical Assistance; Validity
IDENTIFIERS	Education Consolidation Improvement Act Chapter 1; Office of Education; Technical Assistance Centers; *Title I Evaluation and Reporting System

ABSTRACT

The Title I Evaluation and Reporting System (TIERS) was assessed to determine the effectiveness of the TIERS in meeting its statutory objectives. A second purpose was to outline directions for the future of the system in light of enactment of the Educational Consolidation and Improvement Act of 1981. Focus was on assisting the United States Department of Education in conducting both the review of the models in its 1979 regulation and the review of the TIERS requested by the House of Representatives and the Senate in 1978. Generally, local and state educational agencies administering Title I programs are responsible for implementing the TIERS evaluation models, developed by the Office of Education (OE) to implement required standards. To assist these agencies, the OE provides technical assistance in the application of the models. Regular collection of data, via the models, was done at the state and local level. The OE aggregated all data to provide a national overview of the program's effectiveness. The Department of Education commissioned research into the five topics relevant to the current effectiveness and future direction of the TIERS. The 4-month assessment was conducted by a panel of five researchers, each of whom analyzed a separate aspect of the TIERS. Specific topics covered include the validity of TIERS, technical assistance centers, the utility of the TIERS system, changes in evaluation practice at local and state levels and the possible influences of the TIERS, and the Federal Government's perspective concerning the TIERS. Seven data tables and one figure are provided. Six appendices include the Title I evaluation requirements and supplementary text. (TJH)

ED309196

ASSESSMENT OF THE TITLE I EVALUATION AND REPORTING SYSTEM

Elizabeth R. Reisner
NTS Research Corporation

Marvin C. Alkin
University of California at Los Angeles

Robert F. Boruch
Northwestern University

Robert L. Linn
University of Illinois

Jason Millman
Cornell University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

April 1982

Submitted to
U. S. Department of Education
Office of Planning, Budget, and Evaluation
Washington, D.C.

This report was prepared pursuant to the following contracts: 300-82-0008, 300-82-0009, 300-82-0010, 300-82-0011, and 300-82-0012. It does not, however, necessarily reflect the views of the U. S. Department of Education.

The report should be referenced as follows:
Reisner, E. R., Alkin, M. C., Boruch, R. F., Linn, R. L., and Millman, J. Assessment of the Title I Evaluation and Reporting System. Washington, D.C.: U. S. Department of Education, April 1982.

TM013682

BEST COPY AVAILABLE

ASSESSMENT OF THE

TITLE I EVALUATION AND REPORTING SYSTEM

Digest of Conclusions

Main accomplishments of the Title I Evaluation and Reporting System (TIERS) and the associated program of technical assistance:

- Title I data have been provided to Congress in a form specified by legislation.

- Improvements based on TIERS information have been made in Title I project implementation and evaluation practices at state and local levels.

- Through Technical Assistance Centers (TACs), state education agencies (SEAs) and local educational agencies (LEAs) have learned new evaluation approaches and new ways to use evaluation data to improve educational services.

Problems in the current system and implications for change:

- TIERS is currently unstable due to recent changes in federal law.

  -- The U. S. Department of Education (ED) should strongly encourage SEAs and LEAs to continue systematic evaluations under new legislation replacing Title I.

- Current estimates of project impact overstate achievement gains.

  -- LEAs should shift from fall-to-spring testing to annual testing cycles (e.g., fall-to-fall or spring-to-spring).

- Evaluation models have limited value for small districts.

  -- Impact estimates should be reported in terms of a range, reflecting the degree of uncertainty of the estimates.

- TIERS alone is insufficient for meeting local data needs.

  -- TACs should help LEAs reformat evaluation results to be more useful locally. SEAs and LEAs should be encouraged to develop new evaluation strategies to meet their own needs.

- The cost of TAC assistance is high relative to the amount of service provided.

  -- ED should consider options for reducing costs while maintaining the current quality of services.

- States vary in evaluation capabilities.

  -- Differences should be recognized, and services to SEAs should be varied according to SEA needs and interests.

SUMMARY AND CONCLUSIONS

This assessment of the Title I Evaluation and Reporting System (TIERS) was undertaken at the request of the U. S. Department of Education (ED), in order to determine the effectiveness of the system in meeting its statutory objectives. A secondary purpose of the assessment was to outline directions for the future of the system, in light of the enactment of the Educational Consolidation and Improvement Act (ECIA) of 1981. Chapter I of that act replaced Title I of the Elementary and Secondary Education Act (ESEA) and repealed the key authorization of TIERS that had existed since 1974 under Title I.

The 1974 authorization had required the U. S. Office of Education (OE) to "develop and publish standards for evaluation of program or project effectiveness" that would be suitable "for producing data which are comparable on a statewide and nationwide basis" (Section 183 of ESEA Title I). State education agencies (SEAs) and local educational agencies (LEAs) administering Title I programs were responsible for implementing the TIERS evaluation models, developed by OE to implement the required standards. To assist SEAs and LEAs in this task, OE was required to provide technical assistance in the application of the models. Descriptive data on Title I activities and project impact data generated through the use of the models were required to be collected by LEAs and reported to their respective SEAs on a regular basis. SEA reports combining data from many LEAs were then to be submitted to OE, where the SEA data would be aggregated to prepare a national overview of Title I effects. Reports reflecting this national picture were required to be submitted to Congress biennially.

To accomplish the review of TIERS presented here, ED commissioned studies on five topics relevant to the current effectiveness and future direction of TIERS. The five studies are summarized in the second through the sixth chapters of this report, following a detailed introductory chapter. The findings of Chapters Two through Six are discussed briefly here.

Validity of TIERS

In Chapter Two, Robert L. Linn and his co-authors examine the validity of the estimates of project impact generated by the TIERS

models. Factors such as statistical artifacts, relevance and quality of norms, conversion errors, student selection procedures, test administration, and practice effects are explored to assess their effects on the gain scores produced by TIERS. Linn concludes that the small biases produced by several of these factors are in the same direction and are cumulative. These biases are far more severe when estimates are based on fall-to-spring testing cycles, than when they are based on an annual testing cycle. He suggests that much of the bias (as well as considerable test burden) could be avoided by elimination of fall-to-spring testing cycles and by adoption of annual (fall-to-fall or spring-to-spring) cycles. Linn's study confirms the work of other researchers suggesting the difficulty of using the TIERS models to identify relatively small gains in achievement that result from projects involving small numbers of Title I participants.

Services Provided by the TACs

In Chapter Three, Jason Millman examines the current activities and effectiveness of the Title I Technical Assistance Centers (TACs), which are supported by ED to provide help to SEAs in their implementation of TIERS. He finds that, since their establishment in 1976, the TACs have gradually shifted from an emphasis on helping clients implement the TIERS models to an emphasis on improving the quality and utility of information generated by TIERS. Millman finds widespread satisfaction among SEAs with the performance of the TACs. Most SEA respondents see a continuing need for TAC services and want the Centers to be continued. Despite this level of approval, Millman finds substantial variations among the SEAs in their use of the TACs. He also reports that the federally supported services provided by TACs are quite expensive, when measured in terms of costs per hour of TAC service.

Utility of TIERS Information for Improving Educational Practices

In Chapter Four, Marvin C. Alkin and his colleagues describe the utility, or extent to which it is possible to utilize, TIERS and other Title I evaluation data to improve the quality of educational programs. At the LEA level, Alkin finds that evaluations are used in modifying attitudes, recognizing or identifying problems, prompting administrative

or organizational action, and inducing instructional change. At the SEA level, he finds that Title I evaluations are being used by SEAs to assist in the recognition of problems requiring programmatic intervention at the local level and in the recommendation of solutions for those problems. In his study, Alkin has identified six contextual variables affecting evaluation use at the LEA level. Among these variables, the extent to which evaluation data are tailored to the needs and interests of the local community constitutes a primary factor in evaluation use.

Effects of TIERS on State and Local Evaluation Practices

In Chapter Five, Robert F. Boruch and his colleagues review changes in state and local educational evaluation practices that have occurred in connection with TIERS. They find that state and local educators and evaluators are considerably more knowledgeable about issues in estimating educational effectiveness than they were prior to the implementation of TIERS; this awareness often encompasses an understanding of the limits of evaluation, including problems posed by small project size. Boruch and his colleagues also report an increase in the number and quality of studies and analyses initiated by large SEAs and LEAs as part of their efforts to improve educational programs. They find, however, that the federal emphasis on evaluation appears to have had no effect in fostering high quality tests of the costs and benefits of alternative educational approaches. Changes in small LEAs, they conclude, are plausibly attributable to TIERS and its support mechanisms, including TACs and direct assistance from ED and from Title I evaluation units in SEAs. A sizeable portion of all large LEAs need no assistance in implementing TIERS but do use federal initiatives as leverage in improving both the quality of their evaluations and their utilization of evaluation results.

Uses of TIERS Information in Meeting Federal Needs

In Chapter Six, Elizabeth R. Reisner describes the effectiveness of TIERS in meeting the information needs of Congress and the Executive Branch. She finds that the federal level provisions of TIERS have now been fully implemented. With the issuance of the FY 1981 Annual Evaluation Report on Education Programs, ED has fulfilled its obligation to report national Title I data generated through the local use of the TIERS models. Reisner also finds that TIERS is providing data on topics

of interest to persons working on Title I issues in Congress and the Executive Branch; those topics focus primarily on the achievement gains of Title I participants in the basic skills. She reports relatively little effect of TIERS on evaluation and reporting requirements in other federal education programs. She does, however, find reason to expect that Congress will continue to need effectiveness data on federal compensatory programs, despite the change from Title I to Chapter I.

Accomplishments of TIERS

The report suggests that the program of Title I evaluation and technical assistance activities has achieved particular success in three areas. First, the program has been used to provide achievement data to Congress in a form which the Members of Congress themselves specified. Second, the program of evaluation and technical assistance has had a positive effect on the quality of local and state practices in evaluation and program implementation (e.g., needs assessment and student selection in local Title I projects). Third, through the TACs, many local and state educators have been introduced to new evaluation approaches, and attention has been focused on the use of evaluative information for program improvement purposes.

Problems and Implications for Change

Many federal level users of TIERS data, especially those in Congress, erroneously believe that the system is a stable, institutionalized component of state and local programs of evaluation and reporting; this misunderstanding indicates that they are unaware of the uncertainty that currently exists as to whether the TIERS activities will be continued.

- Based on the findings of the report, the authors urge that ED clarify the scope of state and local responsibilities for Chapter I evaluation and reporting. That clarification should strongly encourage the continuation of a cooperative system of evaluation and reporting that will provide information to address objectives of Chapter I program evaluation. If SEAs continue to use a TIERS-like system, several problems existing in the current TIERS system must be addressed. These problems are described below, along with accompanying suggestions for program change.

Due to technical difficulties that are potentially correctable, the estimates of project impact on student achievement are biased in the direction of overstating achievement gain.

- An annual testing cycle (either spring-to-spring or fall-to-fall) would substantially reduce the bias present in the fall-to-spring testing cycle, which is now used most frequently by LEAs. Such a change should be urged by ED.

For smaller-sized districts, which contitute the majority of project sites, the models for measuring project impact on student achievement have limited value and even invite misuse.

- The local level estimates of project impact should be accompanied by an indication of the degree of uncertainty of the estimates. This might take the form of stating estimates as a range of values rather than a single number, which gives a misleading appearance of precision.

TIERS alone is insufficient as a data source for meeting local Chapter I related information needs. (This fact is acknowledged in official TIERS publications, but it is not widely recognized in the field.)

- Chapter I evaluation utility should be improved by continued technical assistance in reformating TIERS-like results to meet LEA information needs. In addition, SEA and LEA evaluation offices should be encouraged to design a variety of local decision-focused evaluation strategies. In particular, locally designed evaluations might provide information on impacts and costs of various instructional materials and processes used within projects.

The cost of technical assistance services to SEAs and LEAs has been high relative to the amount of services provided.

- The delivery of technical assistance should be changed in ways that will decrease the unit cost of the services while maintaining their quality and prominence. The possibilities for changes range from making minor alterations to the present TAC operations to instituting a drastically different

system (e.g., vouchers for purchasing technical assistance services from approved providers).

Finally, states vary considerably in their evaluation capabilities, leading to wide variations in how well evaluations are implemented and reported. In fact, without technical assistance, some SEAs would be unable to implement even minimally acceptable evaluations.

- Given limited resources, the federal government should target technical assistance to the SEAs that have both strong interest in and need to improve their Chapter I evaluation practices. Technical assistance should be provided directly to LEAs in states that lack both interest and capability. SEAs and LEAs that have strong capabilities should be viewed as major Chapter I resources for ideas, approaches, and other products for both evaluation and project improvement.

TABLE OF CONTENTS

DIGEST OF CONCLUSIONS 1

SUMMARY AND CONCLUSIONS..... 11

1. BACKGROUND AND PURPOSE OF THIS ASSESSMENT OF THE TITLE I EVALUATION AND REPORTING SYSTEM..... 1

2. THE VALIDITY OF THE TITLE I EVALUATION AND REPORTING SYSTEM 7

Estimated Project Impact 8
Validity of the Estimates..... 10
Model A..... 10
Cumulative Effects..... 20
Project Level Information..... 22
Conclusions..... 24

3. THE TECHNICAL ASSISTANCE CENTERS..... 27

Background..... 27
Structure, Operations, and Mission of the TACs..... 29
Services..... 30
Costs and Effectiveness of the TACs..... 35
Technical Assistance in the Future..... 39
TAC Characteristics (Plus and Minus)..... 42
Concluding Thoughts..... 43

4. THE UTILITY OF TITLE I EVALUATION..... 46

Previous Research..... 47
Methodology..... 49
Evaluation Activities and Uses — SEA Level..... 51
Evaluation Activities and Uses -- LEA Level..... 54
Factors Leading to Utilization: The Case Studies..... 58
Conclusions..... 63

5. CHANGES IN EVALUATION PRACTICE AT LOCAL AND STATE LEVELS AND THE POSSIBLE INFLUENCES OF TIERS..... 67

The Needs of LEAs, the Needs of Students, and Program Planning..... 67
Achievement Testing and Implementing the Evaluation Models and Reporting System..... 71
Capabilities, Knowledgeability, and Independence of Evaluation Staff..... 77
Going Beyond the Reporting Requirements..... 80

6. THE VIEW OF TIERS FROM WASHINGTON..... 85

Methods of Inquiry..... 86
Federal Implementation of the Legislative Mandate for TIERS..... 87
TIERS Success as an Evaluation System..... 96
TIERS as a Model for Other Program Evaluation and Reporting Systems.....102
Future Federal Needs for TIERS..... 105

REFERENCES.....110

GLOSSARY OF ABBREVIATIONS.....113

APPENDIXES.....114

A. Title I Evaluation Requirements, As Contained in the Education Amendments of 1978 114

B. Chronology of Communications Between O.' and Congress on TIERS.....117

C. Evaluation Requirements, As Contained in the Education Consolidation and Improvement Act.....124

D. Planning and Advisory Group for the Assessment of TIERS.....125

E. SEAs and LEAs in Which Interviews Were Conducted for This Study.....126

F. Federal Personnel Interviewed for This Study.....127

LIST OF TABLES AND FIGURES

Figure 2.1	Pretest and posttest NCE averages in reading achievement for fall-spring and annual testing cycles in grades 2 through 6	21
Table 3.1	The magnitude of TAC field service	31
Table 3.2	Distribution of TAC field services by states	33
Table 3.3	Cost of client services	37
Table 4.1	Instances of evaluation use in exemplary districts	56
Table 4.2	Case study information	58
Table 5.1	Status of model implementation by LEAs in Region I	74
Table 5.2	Percent of LEAs implementing models	75

CHAPTER ONE

Background and Purpose of This Assessment of the Title I Evaluation and Reporting System

In 1974 Congress added a new set of evaluation requirements to Title I of the Elementary and Secondary Education Act (ESEA). The requirements, as set forth in Section 151 of the Education Amendments of 1974, charged the Commissioner of Education with the development and implementation of "standards for evaluation of program or project effectiveness in achieving the objectives" of Title I. The objectives to be measured were to be those guiding Title I from the time of its enactment in 1965. Evaluation would assess the success of Title I in expanding and improving educational programs that meet "the special educational needs of educationally deprived children" attending schools in "areas with concentrations of children from low-income families" (Section 101 of the act).

The standards required by Congress in 1974 were to be implemented through an evaluation system containing several distinct features, including "models for evaluations of all programs conducted" under Title I, technical assistance to state education agencies (SEAs) "to enable them to assist local educational agencies (LEAs) in the development and application of a systematic evaluation of programs in accordance with the models," an annual evaluation report to Congress (later made biennial), and a dissemination program providing information to SEAs and LEAs on exemplary Title I projects.

This congressionally mandated system reflected an ambitious view of the potential role of local evaluation in the operation of the Title I program. Evaluation at the local project level was seen by Congress as a key vehicle for assessing the effectiveness of the Title I program nationwide and also as a means for generating information locally that could be used as the basis for local and state decisions on project improvement. This dual focus was discussed in the 1974 House report on the reauthorization of Title I (Committee on Education and Labor, 1974, p. 20):

> What is needed in terms of a national evaluation is for the Office of Education to specify the critical information which States should provide, including data on the impact of the program on educational achievement. Further, the States should be required to report such information in a standardized format to facilitate the compilation of a report of national scope on Title I program effectiveness, and to establish a base for formation of program policy direction. In-depth technical assistance should be provided to the State education agencies in the area of evaluation, including such items as the information required, format, procedures, and analyses. . . . Local educational agencies' project evaluations should yield objective achievement data on which local officials can base subsequent project modification or which can be used to identify exemplary projects.

The House committee stated its belief that these objectives could be achieved within the then-existing framework of the law and required no special amendments. Nevertheless, during final floor debate over other provisions of Title I, an amendment, which later became Section 151, was offered by Representative Victor Veyzey (R.-Cal.). The amendment was approved by the full House without preliminary review by the Education and Labor Committee and later accepted without report language by the House-Senate conference committee on the ESEA amendments. Thus enacted, Section 151 laid the basis for what is now known as the Title I Evaluation and Reporting System (TIERS). TIERS is made up of the Title I evaluation and reporting requirements that are mandatory in nature, which include adherence to the technical standards for Title I evaluations, use of the Title I evaluation models, and periodic reporting of Title I information by LEAs to SEAs and then to the federal government.

Although the U.S. Office of Education (OE) took prompt steps to begin implementation of TIERS, the system was not installed before the next legislative review of Title I, which occurred in 1978. As a result of that review, several minor changes were made to the 1974 provisions, mainly requiring OE to consult with state and local evaluators in the design of the Title I evaluation system. In addition, Section 151 became Section 183 of the Education Amendments of 1978. Section 183 and the requirements for the conduct of local evaluations contained in Section 124(g) constituted the key evaluation requirements for Title I for the succeeding three years. Relevant statutory provisions for TIERS and other aspects of Title I evaluation are reprinted in Appendix

A of this report. Appendix B presents a chronology of communications between Congress and the OE regarding the development of TIERS.

Despite the relatively minor changes made by Congress in 1978, the House and Senate committee reports on Title I leveled serious criticisms at OE's implementation of the 1974 evaluation requirements (Committee on Education and Labor, 1978, pp. 50-53, and Committee on Human Resources, 1978, pp. 27-30). Both reports (in identical language) requested OE to review its implementation of the Title I evaluation procedures:

> As GAO [the General Accounting Office] recommended, the Committee believes that OE should assess whether State and/or local evaluation reports for Titles I and VII can be improved so that they supply officials at Federal, local, and/or State levels with the reliable program information they need for decisionmaking. This includes assessing the adequacy of the Title I evaluation models and related data. If OE determines that it is unrealistic to expend resources on improving these programs' State and local evaluation reporting systems based on aggregated data, then it should take the needed steps to adopt more feasible and effective approaches. This should include eliminating unwarranted reporting requirements and if necessary proposing to the Congress any legislative changes needed to accomplish this. . . .
>
> In connection with the assessment recommended above, GAO also recommended that the Secretary of HEW [Health, Education and Welfare] direct OE to review the types of State and/or local program evaluation information collected (or planned to be collected) on programs authorized by Titles I and VII of ESEA. The Committee believes that this recommendation should be implemented and the review should include an assessment of the information's usefulness at each level and should determine:
>
> --Whether it is realistic to attempt to serve Federal, State, and local levels with aggregated data based on local agency evaluation reports.
> --How the information needs at the local, State, and Federal levels can best be met.
> --Whether unnecessary duplication exists or will exist in meeting Federal information requirements through the State and local reporting systems as well as through OE national evaluations of these programs, and if so, how it should be eliminated.
>
> During this review process to define the evaluation information needed at State and local levels, OE should seek the views and cooperation of the State and local officials who are intended to use the results.

The uncertainty expressed here about the utility of the Title I evaluation system was reflected again in 1979 in the preamble to the final regulations implementing TIERS. The preamble prepared by OE states, in part:

> The required evaluation models represent an improvement over the practices and procedures of many locally conducted Title I evaluations. Technical questions remain, however, concerning such issues as the extent to which the different models yield comparable data. These issues are currently under investigation by OE. As further technical analysis leads to refinements in either the models or the process for reporting evaluation results, revisions in the regulations may be needed. The Commissioner intends to reconsider, and if necessary, revise the regulations after a three-year period, based on information available at that time. Representatives of SEAs and LEAs will be invited to participate in this review.

The report presented in the following chapters is intended to assist the U.S. Department of Education (ED) in carrying out both the review of the models promised in its 1979 regulations and the review of TIERS requested by the House and Senate in 1978.

These purposes of the report are, however, substantially complicated by the enactment in 1981 of the Education Consolidation and Improvement Act (ECIA). Chapter I of ECIA amended Title I extensively in an effort to "eliminate burdensome, unnecessary, and unproductive paperwork and free the schools of unnecessary Federal supervision, direction, and control" (Section 552 of ECIA). ECIA repealed Section 183 of Title I, which had been the key authorization of TIERS. It also reduced the evaluation requirements imposed on local Title I projects. As a result, the remaining evaluation requirements in what is now called Chapter I include only a provision requiring SEA recordkeeping and reports to the ED Secretary and a provision requiring local evaluations of basic skills achievement. Chapter III of ECIA authorizes ED to provide technical assistance to SEAs and LEAs for certain purposes, including instructional improvement. These provisions of ECIA are reprinted in Appendix C.

Seen in the context of the enactment of Chapter I, the assessment of TIERS presented in this report can be viewed as having potential uses that are both retrospective and prospective in nature. The assessment reports on the implementation of TIERS and the effects of that implementation. Its findings should permit a number of different kinds of judgments to be

made. Included among them may well be decisions as to where TIERS should go from here.

The assessment was conducted over a period of four months by a panel of five researchers, each of whom looked at a separate aspect of the Title I evaluation and reporting system. Each panelist was responsible for drafting a chapter describing the issues, methods of inquiry, and results of his or her part of the overall assessment. All five panelists collaborated in the preparation of the Summary and Conclusions, presented first in the report. For purposes of coordination, one member of the panel, Elizabeth R. Reisner, served as its chair.

As noted, the subsequent chapters of this report address five topics. In the first of these chapters Robert L. Linn, Professor of Education and Psychology at the University of Illinois, describes the technical feasibility of the TIERS models as a method for producing valid evaluative data at local, state, and national levels. Jason Millman, Professor of Educational Research Methodology at Cornell University, examines the current activities and effectiveness of the Title I Technical Assistance Centers (TACs), which are supported by ED to provide help to SEAs in their implementation of TIERS. Marvin C. Alkin, Professor of Education at the University of California at Los Angeles, describes the usefulness of TIERS (and other evaluation data) in improving the quality of local and state educational programs. Robert F. Boruch, Professor of Psychology and Education at Northwestern University, reviews changes in state and local educational evaluation practices that have occurred as a result of TIERS. Elizabeth R. Reisner, Senior Analyst at NTS Research Corporation, describes the effectiveness of TIERS in meeting the information needs of Congress and the Executive Branch.

The panel wishes to acknowledge the assistance provided by the planning and advisory group assembled by ED to launch this assessment. The names of the members of that group are presented in Appendix D. Many other individuals at local, state, and federal levels generously donated their time, their recollections, and their insights to the interviewers who collected data for this report. Interview respondents at state and local levels included educational administrators, evaluators, teachers,

and parents. It is not possible to thank them all individually. Appendix E lists the SEAs participating in this study and the numbers of LEAs in each state that were contacted for the study. Appendix F lists the federal-level personnel who were interviewed for the study.

The panel also wishes to offer special thanks to Robert M. Stonehill of the ED Office of Planning, Budget, and Evaluation and his colleagues Judith Anderson and James English for their assistance in this project.

CHAPTER TWO

The Validity Of The Title I Evaluation and Reporting System

Based on results of tests administered to 298,833 third grade students in the fall of 1979 and again in the spring of 1980, the estimated gain in reading achievement due to Title I participation for the nation as a whole is 7.3 Normal Curve Equivalents (NCEs). The corresponding gain in mathematics achievement of 2,464 fourth grade Title I participants in the state of Kentucky was 9.4 NCEs, while 36 second graders in reading project number 0115A showed a gain in reading achievement of 12.1 NCEs. These are but three among thousands of numbers produced by TIERS. These numbers are intended to answer the question: "How much more did students learn by participating in the Title I project than they would have learned without it?" (Tallmadge & Wood, 1976, p. 2). The purpose of this chapter is to evaluate the trustworthiness of the answers to this basic question derived from TIERS.

The focus of this chapter is on the technical adequacy of TIERS. The guiding question for our analysis was: "How valid are the estimates of the impact of Title I on student achievement that are derived from TIERS?" This seemingly straightforward question has many facets and is subject to several interpretations. For example, the system is designed to provide information about the impact of Title I at the local, state, and national levels. The importance of factors that affect the validity of estimates and the conclusions that are based on them may differ substantially at different levels of aggregation.

Questions of validity may be interpreted narrowly (e.g., are the scores reported for a particular test accurate?) or broadly (e.g., is

---

The authors of this chapter are Robert L. Linn, Stephen B. Dunbar, Delwyn L. Harnisch, and C. Nicholas Hastings. We thank Leigh Burstein and Richard Jaeger for their advice on the project and critique of drafts. We also thank the many people involved with Title I evaluations at local and state levels, at Technical Assistance Centers, and at RMC Research Corporation who provided us with information and advice.

Title I effective in improving the achievement of participating students?). Although the broader questions are generally more difficult to answer and require more interpretation, they are also more important and will guide much of our discussion.

Estimated Project Impact

TIERS is designed to provide information on several aspects of Title I projects. A good deal of the information is descriptive in nature. For example, TIERS provides data on the number of students participating in Title I projects, on the extent of parental involvement in school advisory councils, on Title I staff, and on the type of services provided to participating students. The primary focus, however, is on student achievement in the basic skills: reading, language arts, and mathematics. Estimates of impact on student achievement are obtained at the project level, which might, for example, involve a half-dozen students who are pulled out of their regular classes for three or four hours a week and given special instruction in reading by a Title I teacher. On the other hand, a "project" may have several hundred participating students who receive instruction from dozens of different teachers at a variety of school buildings within a district. Thus, the term "project" has rather diverse meanings.

Data from individual projects are reported to an LEA which, in turn, analyzes, summarizes, and reports them to the SEA. Additional analyses and summarizations occur at the state level in preparing the state Title I report and in transmitting the information to the federal level. The estimated gains in achievement listed in the opening paragraph of this chapter are three of thousands of estimates that result from this process.

What does it mean to say that third grade Title I participants gained 7.3 NCEs from fall to spring? This is an overall estimate of the typical or average impact of Title I, which is obtained by comparing the achievement test results that are actually obtained in the spring with the "no-project expectation." The latter value is based on the performance of the participants when tested in the fall in comparison to normative data reported by publishers of standardized tests. A value of zero is intended to be

indicative of no positive impact of Title I beyond what would have resulted without it. Any positive number is supposed to indicate a beneficial effect of participation in Title I.

The estimated gains are reported using a scale called Normal Curve Equivalents, or NCEs. This scale ranges from 1 to 99 and is based on normative data provided by test publishers. NCEs and percentile ranks have the same numerical value at 1, 50, and 99. They differ at intermediate values, but there is a unique percentile rank associated with each NCE. For example, an NCE of 10 corresponds to a percentile rank of 23. In other words, 23 percent of the norm group scores below an NCE of 10.

The gain of 7.3 NCEs may be made more concrete by reference to the actual fall and spring results. In the fall, the mean reading achievement of third grade Title I participants was 28.6 NCEs and by spring it was 35.9, for a gain of 7.3. In more familiar terms, the average Title I third grade student scored higher than only 15.5 percent of all third graders in the test publisher's normative samples when tested in the fall, but when retested in the spring the average Title I student scored higher than 25.1 percent of all such students. 1/ In the absence of Title I, it is assumed that the standing of participants would have remained constant relative to the national norm.

The gain of 7.3 NCEs (or from the 15th to the 25th percentile) sounds good, but there are many questions that need answers before it can be accepted. This is clearly recognized in ED's FY 1981 Annual Evaluation Report on Education Programs, from which the number was obtained. In that report it is noted that the results "must be viewed with great caution, since these data: (1) are based on scores aggregated across many different tests, of varying quality and appropriateness, and administered under locally controlled conditions; (2) in some situations were obtained in States or districts implementing new evaluation procedures for the first time; and (3) in some situations quality control procedures may not have been adequately implemented" (p. 30). The remainder of this chapter is devoted to evaluating the seriousness of these and other threats to the validity of the estimates of impact and to suggesting, where possible, ways of alleviating problems.

1/ The 15.5 percent and 25.1 percent are derived from the fact that an NCE of 28.6 equals a percentile rank of 15.5 and an NCE of 35.9 equals a percentile rank of 25.1.

Validity of the Estimates

There are numerous factors, some already mentioned, that may affect the validity of the estimates of impact on student achievement. Questions may be asked about comparability of results based on the use of different evaluation models, of different testing patterns, and of different tests. Testing conditions, sample attrition, and quality control are but a few of the technical issues that are of potential concern. Moreover, the tenability of the basic assumptions used to derive the "no-project expectation," which is the fundamental benchmark for any of the estimates, is subject to serious question. We have attempted to deal with the most critical concerns in as nontechnical a way as possible. More technical discussion, where necessary, has been reserved for a separate report.

Three basic models were developed and refined by RMC Research Corporation for inclusion in TIERS. Variations are allowed in each of the models, and, with appropriate approval, locally defined alternatives may also be used. Practically speaking, however, there is only Model A, the norm-referenced model, since use of the other recommended models, or of approved exceptions, is extremely rare, accounting for only about 1 percent of the project evaluations. Consequently, we have focused almost exclusively on Model A. 2/ In discussing Model A and its implementation, we focus on the cumulative effects of statistical artifacts, unrepresentative norms, conversion errors, student selection, test administration procedures, attrition, and practice effects on the validity of estimates derived from TIERS.

Model A

Test publisher norms provide the basis for the "no-project expectation" in Model A. It is simply assumed that, in the absence of participation in a Title I project, participants would maintain, on the average, a constant position relative to the norm. For example, if on the average,

---

2/ Although the 1976 User's Guide stated a preference for Models B and C over Model A, subsequent research (e.g., Stewart, Note 1; Tallmadge & Wood, 1981) led to a reassessment of Model C, which was found to produce erratic and biased estimates. Model B was used in only about one-tenth of one percent of the projects in 1979-80, and when used, practical compromises required to implement it largely, if not entirely, eliminate the advantages of Model B (see Tallmadge, Note 2).

participants scored higher before the project than 20 percent of the nation's students in the same grade and at the same time of year, then they would be expected to score higher than 20 percent of all students following the project. Since students in the norm group improve their skills and score higher at the time of the second test than at the time of the first test, the project participants would have to get more answers right on the second testing in order to maintain a constant percentile rank. To show a positive estimated impact, the Title I group would have to gain faster than the norm.

The assumption of constant relative standing in the absence of a project, commonly called the equipercentile assumption, has been questioned by a number of authors. The assumption has been challenged because it is inconsistent with established statistical theory, 3/ because the target group in Title I evaluations may not be well represented in the norm group, and because the norms are obtained under testing conditions that may differ from those used in the Title I evaluation.

Statistical artifacts. The models developed by RMC Research Corporation were designed to minimize a statistical artifact, known as the regression effect, that is frequently a serious problem in evaluation studies. If properly implemented, Model A procedures clearly reduce the bias in estimates of project impact due to the regression effect. A careful analysis shows, however, that the regression effect is reduced but not completely eliminated by use of Model A. In other words, based on statistical theory, there are reasons to expect the estimated gains produced by Model A to be somewhat too large (see Glass, Note 3, or Linn, 1980, for discussions of these reasons). Research conducted primarily by staff of the RMC Research Corporation (see, for example, Roberts, 1980; Tallmadge, Note 2; and Wood, Note 4) has convincingly shown that, although there is a bias in the estimates of gain in the direction that is predicted from statistical theory, the bias is not large. As summarized by Tallmadge (Note 2), "The bias appears to be positive

---

3/ A well known, albeit frequently misunderstood, statistical phenomenon known as "the regression effect" would lead to an expectation that is contrary to the equipercentile assumption.

and on the order of 1 NCE when typical Title I groups are examined..." (p. 30 of typescript). In other words, the estimated gain is artifically inflated due to a statistical artifact by about 1 NCE. An adjustment for this small, but systematic, bias in the estimated gain would be simple and possibly worthwhile. As will be seen, however, it is not the most serious threat to validity.

Relevance and quality of norms. The validity of the TIERS results depends on the quality of normative data and the comparability of the testing for normative and for Title I evaluation purposes. For an evaluation, only students who are tested before and after (the pretest and the posttest) are included in the estimate of the impact of a Title I project. Publishers norms relevant for the pretest and for the posttest, however, typically include students who are tested one time but not the other. In the case of a spring-to-spring testing pattern, the norms are based on entirely different students. For a fall-to-spring testing pattern the norms are usually based on a mixture of students tested on both occasions and on only one of them. Seventy percent of the school districts included in the fall norms of one test publisher, for example, also participated in the spring norming. Another publisher had all the districts participate on both occasions. A third publisher reported overlap of about 20 percent of the students.

Results for students tested both in the fall and in the spring exclude students who are available for testing only in the fall or only in the spring. There is evidence that the norms based on students tested both fall and spring (i.e., longitudinal norms) are somewhat different than those for all students tested on either occasion (i.e., cross-sectional norms). Results reported by Kaskowitz and Norwood (Note 5) for one major test publisher suggest that the practice of comparing the longitudinal data required for Title I evaluation to cross-sectional norms provided by a test publisher results in a small positive bias of roughly 1 NCE on the average. Although this result is based on only one test and must be viewed with some caution, it still suggests that the use of cross-sectional norms causes a small bias in the same direction and

of the same order of magnitude as the bias due to statistical regression discussed above. These biases work to inflate the estimated gains, and the effects can be expected to be cumulative. Together they may mean that the gains are overstated by roughly 2 NCEs.

In addition to the question of appropriateness of cross-sectional norms for longitudinal evaluation data, questions have been raised about the relevance and representativeness of the norms. Test publishers must rely on voluntary cooperation of school districts to collect normative data. Districts are scientifically selected by publishers to yield norms that are representative of the nation as a whole. But many of the selected districts decline to participate. Indeed, according to Baglin (1982), invited districts are more apt to refuse than to agree to participate. Consequently, the "purportedly random 'national' norming samples are not at all random. . . . In other words, each test is normed on a nonrepresentative, nonrandom cross section of the population" (p. 105). Not only are participation rates low, but they have apparently declined substantially from what they were in the past.

The impact of lack of representativeness of publisher norms on the validity of TIERS is problematic. Representativeness would clearly be crucial if the focus in TIERS was on the students' level of achievement at one point in time. The focus in TIERS, however, is on gains relative to the norm rather than on assessing achievement status and even nonrepresentative norms may provide a useful benchmark for this purpose. On the other hand, it is true that different normative samples can produce different estimates of NCE gains from a given set of test results (see, for example, see S. J. Roberts, Note 6). What remains unclear is whether the differences due to nonrepresentativeness are consistent across publishers, and, if so, whether the bias this introduces inflates or deflates the estimates of gain. Thus, lack of representativeness increases uncertainty regarding validity but does not necessarily bias the estimates.

Conversion errors. In order to obtain NCE scores, tests must first be scored by counting the number of correct answers. The result is known as a "raw score," which must be transformed to NCEs by use of normative tables provided by publishers. Often the conversion is a two step process. Raw scores are entered into a table to find a corresponding percentile

rank, and percentile ranks are then converted to NCEs by another table look-up.

States and districts vary in the degree to which this conversion process is automated. In some instances all of the conversion is done by machine. In other instances, Title I teachers or other local staff score the tests and convert the raw scores to percentile ranks. The final conversion to NCEs is then done more centrally at the district or state level.

A certain amount of human error in recording scores and converting raw scores to percentiles is bound to occur. Studies (e.g., Elman, Note 7; Finley, Note 8) provide evidence that it is indeed the case that project level report forms contain a substantial number of errors of various kinds. Moreover, Elman's results indicate that conversion errors inflate the estimated gains much more frequently than they deflate them, creating a positive bias somewhat larger than 1 NCE. Again, this bias is in the same direction, toward inflated gains, as the two previously identified biases.

Cumulatively, the three systematic sources of bias considered so far amount to roughly 3 NCEs. The third source of bias only applies to districts and states relying on manual conversion of scores at the local level, however. Uniform adoption of currently available computer-based conversion systems could effectively remove this source of bias.

Student selection. In the majority of the project evaluations for TIERS, a "pretest" is administered in the fall and is used to derive the "no-project expectation" against which the "posttest" scores obtained the following spring are judged to determine gain. The standard version of Model A, which is used in almost all cases to estimate the gains, requires that "pretest scores...play no part in the selection process because this will introduce a bias..." (Tallmadge & Wood, 1981, p. 29, emphasis in the original). The developers of Model A had sound statistical reasons for the above requirement, and the statement that student selection on the pretest will create bias is quite right. 4/ If this model

---

4/ There is a variation of Model A in which selection of program participants is based on the pretest. This variation has much to recommend it and will be discussed later. This variation has not been used much, if at all, to date and requires special analytical procedures.

requirement is violated, the estimated gains are artifically inflated. The magnitude of the bias depends on the use of the pretest scores for selection, but the direction of the bias is once again toward an overstatement of gains.

Cautions against the use of the pretest for selection of project participants may have been relatively effective in preventing a major bias due to this misapplication of the model. Based on interviews, however, we concluded that some subtle violations of the stricture against any use of pretest scores occur. Intending to comply with the guidelines, students may be selected for the Title I project prior to administration of the pretest. This may be done appropriately using previous test results, teacher judgments, or a combination of the two. So far, so good, but then the pretest is administered and a few of the selected students do better on it than was expected. It seems intuitively reasonable to use this new information and reverse the initial decision for those students, and, indeed, such selection out of Title I on the basis of pretest scores does occur.

Although using pretest scores in this way may be educationally sound, it introduces a bias that tends to make the estimated gains too large. In the words of one of the evaluators we interviewed, there is a "selection-evaluation dilemma." The goal of doing the best possible job of selecting students from an educational point of view may sometimes conflict with what is needed for the most valid evaluation. We do not have, from available studies, a good basis for estimating the extent of the problem caused by partial selection on the pretest. However, it is one more source of bias in the direction of inflating gain estimates, and while the bias due to this source may be small it adds to the cumulative bias in the direction of overestimating the impact of Title I.

Attrition by "graduation." Students who change schools between the time of the pretest and the posttest are not included in the estimation of gains. As stated in the section on norms, the loss of these students tends to introduce a positive bias in the estimates. However, students who are taken out of the Title I project may also be lost for purposes of the evaluation. If students are taken out of the Title I project

because of their high achievement but not included in the test results, then the gains for the project would be understated. It is important that test results for such midyear "graduates" of the project be obtained and included in the estimates. We lack evidence regarding the extent to which this is done. It could be a source of negative bias in contrast to all the other biases so far identified.

Administration conditions. When tests are normed, they are generally administered to intact classrooms and there is little riding on the outcome for teachers or students. In Title I evaluations on the other hand, the tests are often administered to small groups (e.g., to a group of students that is pulled out of their regular classroom for special instruction by a Title I teacher). There is also a reason to want test scores to be higher in the spring than the fall. One need not invoke a conspiracy theory or accuse teachers of intentionally cheating by altering time limits or giving answers to students to conclude that administrative conditions tend to favor positive results. Such flagrant practices may occur, but we suspect that they are extremely rare, causing no serious distortion of the overall results.

More subtle variations in test administration conditions are apt to be more common and we believe these variations contribute some unknown amount to the cumulative positive bias. One of the people we interviewed, a person with considerable experience with testing and program evaluation, told of an experience he had when working on a testing program where a test administrator complained that the time limits for the test were too stringent, leading to failures to complete the test and lower scores. The "problem" disappeared when the students were warned before the test that they would have a limited amount of time and the time remaining was periodically posted at the front of the room during the test. Such "minor" variations in administration conditions or the presence or absence of pep talks stressing the importance of an exam can affect the results. When there is pressure to achieve a particular result, higher spring than fall scores, the effect even with the best of intentions is apt to be in the direction of producing the desired result. Studies in other contexts

of experimenter bias (e.g., Rosenthal, 1966) support this general conclusion.

Administration dates. The rules for implementing Model A call for the administration of tests within two weeks of the test publisher's norming date. This rule is intended to reduce bias in the estimates caused by using norms obtained at one time of the school year as the benchmark for judging test results of a test administered at a different time of year. By administering the pretest two weeks before the fall norming date and the posttest two weeks after the spring testing date, however, it is still possible to have four more weeks of instruction for project participants between test administrations than was available between the normative dates. Following the above testing pattern would tend to inflate the estimates of gain. Recognizing this potential, some states have added a requirement that the deviations of the test administration dates from the norming dates in the fall and spring be considered together so that the difference in the fall-to-spring interval is no more than 14 days. Such a requirement is potentially useful for reducing the bias. As is discussed below, however, the bigger need is to switch from the predominant fall-to-spring testing cycle to an annual testing cycle.

Practice effects, teaching to the test, and learning. Evidence provided by A.O.H. Roberts (Note 9) based on a careful review of the literature is quite consistent and supports several conclusions regarding practice effects, "test-wiseness," and the effects of coaching students for tests.

- Practice at taking tests tends to yield score increases at all grade levels.
- The gains are larger at the early grades, especially grades 1 and 2 and kindergarten, than later grades.
- Multiple test exposures yield additional, but rapidly diminishing, returns.
- Coaching or explicit instruction in test taking skills can have sizeable effects, sometimes on the order of magnitude of 5 NCEs.
- Test publisher norms reflect increasing test-wiseness of students but cannot compensate for variations in the typical frequencies with which students are tested.

It is important, albeit often very difficult, to distinguish between test-specific learning and learning that is more generalizable, with test performance being just one of a variety of possible indicators. The common concern of a Title I project presumably is not with test performance, per se, but only inasmuch as the performance is indicative of the students' skills, knowledge, and understanding. The dilemma for evaluators and for those who would rely upon test results produced in an evaluation was well summarized almost 20 years ago by Cronbach (1963), who argued that:

> The demand that tests be closely matched to the aims of a course reflects awareness that examinations of the usual sort "determine what is taught." If questions are known in advance, students give more attention to learning their answers than to learning other aspects of the course. This is not necessarily detrimental. Whenever it is critically important to master certain content, the knowledge that it will be tested produces a desirable concentration of effort. On the other hand, learning the answer to a set of questions is by no means the same as acquiring understanding of whatever topic that question represents (p. 681).

One of the features of TIERS has the potential of being both a strength and a weakness in regard to the dilemma articulated by Cronbach. The flexibility in the choice of a test that is built into TIERS enables districts to select a test that is judged to be most appropriate for the particular program. The User's Guide stresses the importance of selecting an appropriate test, noting that "the content of the test must match the content of the instructional program being evaluated" (Tallmadge & Wood, 1981, p. 74). This is sensible advice, because it is clear that "the content of the achievement test used to evaluate a program is critical to the resulting estimate of its effectiveness" (Jaeger, Note 10). but several questions need consideration in interpreting the results. For example:

- If the content of the test and the instruction are better matched in the project evaluation than in the test publishers' norming studies, do the norms provide a fair comparison?
- Are the results generalizable, or are the students "learning the answers" to a specific set of questions rather than "acquiring understanding"?

These are important questions, but they are not ones that have been addressed by the host of studies of the technical quality of TIERS that have been undertaken. Hence, we lack a solid basis for estimating the effects of test specific instruction on the estimation and interpretation of project impact. Indirect evidence from several sources, however, suggests that some inflation of gains is to be expected as the result of learning that is specific to the test.

In addition, apparently trivial changes in the way a question is asked can influence results. Whether addition problems are presented horizontally or vertically, for example, influences scores. More importantly, the scores are affected by the match between the way the problems are presented in the instructional materials and on the test (Alderman, Swinton & Braswell, 1979; Glass, Note 3). Other researchers (e.g., Leinhardt & Seewald, 1981) have reinforced the belief that the match between what is taught and what is tested is vitally important. There is also evidence that superficially similar tests vary greatly in the specifics of content coverage (e.g., Hoepfner, 1978). Careful test selection and/or adjustments in the instructional materials to improve the match provides a project with a net advantage in comparison to the norms against which the gains are judged but leaves doubts about the effectiveness of the project in terms of generalizations beyond the specific test (see House, 1979, pp. 409-411).

Limitations of the Mean. The entire focus of TIERS is on the average gain. While this focus is understandable, changes in means cannot be expected to tell the whole story. An average gain of five points, for example, could result from all participants gaining about five points, from the initially highest scoring participants gaining a lot while the initially lowest scoring participants gain essentially nothing, or from a variety of other patterns. Hence, it would be desirable to have information not only about the changes in the means but on the changes in proportion of participants scoring below specified NCE scores. The latter information could be useful in judging how effective Title I is in raising the floor of the achievement distribution.

Cumulative Effects

A number of threats to the validity of the TIERS estimates of impact have been briefly reviewed. In some instances (e.g., attrition or the lack of representativeness in the norms), threats increase uncertainty regarding the interpretation of the results but do not necessarily bias the estimated gains in one direction or the other. This is also true of a variety of other threats to validity (e.g., floor or ceiling effects caused by tests that are too hard or too easy, and errors in test equating) that were not discussed due to space limitations. 5/

In most cases, however, each threat to validity was found to lead to a small but systematic bias that causes the impact of the program to be overestimated. No single source of bias was large enough, by itself, to invalidate the results. But the effects are cumulative and in the same direction. Together, they probably account for most of the relatively large estimated gains obtained for the bulk of the 1979-80 results which were based on a fall-to-spring testing pattern.

This conclusion is consistent not only with the above analysis of individual threats to validity of the estimates but with other sources of evidence that support the conclusion that the effect of Title I on student achievement is positive, but quite small -- much smaller than the fall-to-spring results would suggest.

Two major national studies, the "Compensatory Reading Study" (Trismen, Waller, & Wilder, Note 11) and the "Sustaining Effects Study" (Wang, Bear, Conklin, & Hoepfner, Note 12), support the conclusion that compensatory education services provided through Title I result in small increases in student achievement in reading and/or mathematics. The order of magnitude of the effect is closer to 1 or 2 NCEs in these studies than the 6 to 9 NCEs estimated from TIERS for the fall-to-spring testing pattern. In other words, rather than estimating that Title I services move a typical third grader from the 16th to the 25th percentile nationally between fall and the following spring, as would follow from TIERS, the gain is more nearly from the 16th to the 17th or 18th percentile.

5/ These issues are discussed in a variety of technical reports that are available through the Title I Evaluation Clearinghouse located at the American Institutes for Research in Palo Alto, Cal.

Corroborative evidence that the fall-to-spring TIERS estimates are artificially large is provided by other data from TIERS itself. Although fall and spring testing is the predominant pattern, used for roughly three-fourths of the students, there is still a sizeable number of students (roughly 100,000 per grade in grades 2 through 6), who are tested on an annual testing cycle. As can be seen in Figure 2.1, the annual testing cycle results yield estimated gains that are considerably smaller than those obtained from fall-to-spring. The annual testing cycle results are also much more consistent with previous findings from national studies.

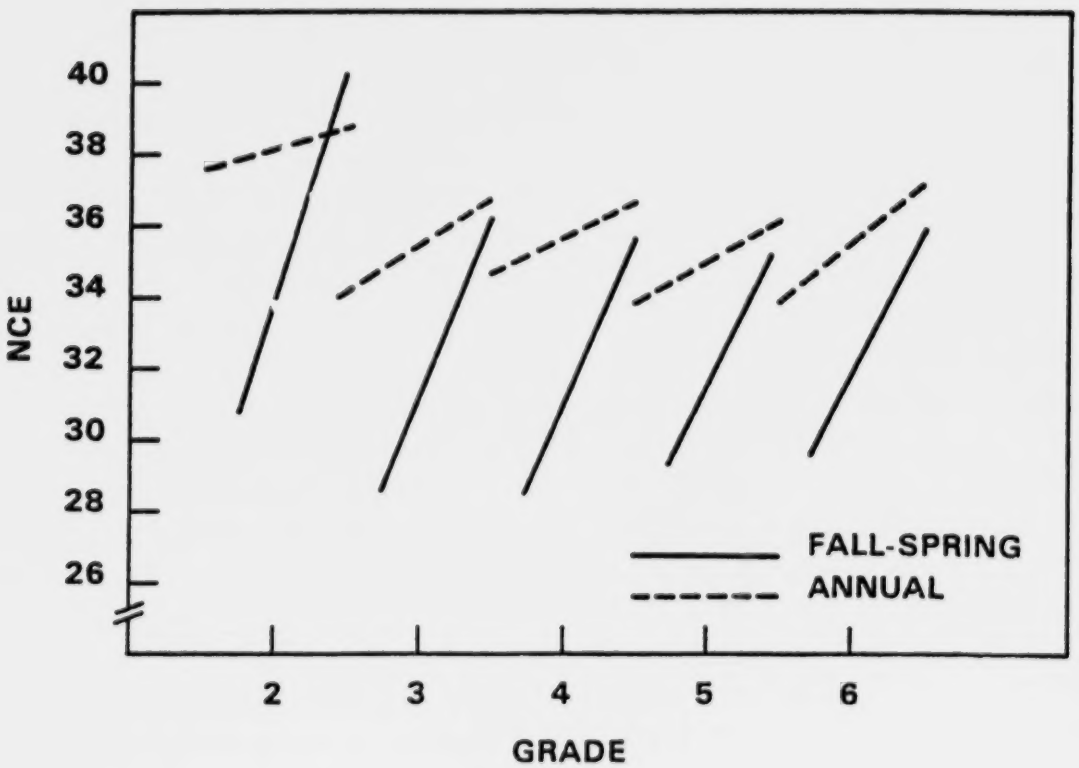

Figure 2.1. Pretest and posttest NCE averages in reading achievement for fall-spring and annual testing cycles in grades 2 through 6 (based on FY 1981 Annual Evaluation Report of Education Programs).

Based on the above analyses, we conclude that the difference in the two results is due to biases in the fall-to-spring estimates. Several of the sources of bias that we have reviewed affect only the fall-to-spring results or are apt to have a larger effect on those results than on results for an annual testing cycle. In an annual testing cycle, one year's post-test may be the next year's pretest. Thus, actions which artificially inflate the scores to yield higher gains for one year lead to a higher starting position and thereby lower gains for the next year.

Even the estimates based on an annual testing cycle are apt to have a small positive bias since some of the factors leading to bias exist for either testing cycle. For example, spring-to-spring results still are subject to small biases due to statistical regression and the use of cross-sectional norms to judge longitudinal data. Hence, an estimated impact of two NCEs is apt to be, if anything, slightly on the high side. 6/

Project Level Information

TIERS was developed in response to a congressional mandate. Hence, it is not at all surprising that the creation of nationally aggregated results was the primary focus in the development and in many of the studies of the technical quality of TIERS. The system is also intended, however, to serve local and state evaluation needs and, through the activities of the TACs, considerable effort has gone into improving the quality of state and local level evaluations.

Although efforts to improve the quality of state and local level evaluations of Title I have probably had many effects, 7/ it is not our

---

6/ Even for the annual testing cycle, the starting performance in each grade is always below the ending performance of the previous grade. While this would seem to suggest that the gains are artificial and there is no real impact, the estimates are based on different groups of students and fewer students participate in higher grades. Thus, the overall trend may paint an overly pessimistic picture.

7/ Several of the people we talked to were of the opinion that activities related to TIERS had greatly improved the understanding of tests and their use in evaluation at the state and local levels. These and a variety of other possible effects of TIERS on evaluation practice are discussed in Chapter Five of this volume.

purpose to try to assess them. Our focus is much narrower. We want to know how trustworthy the estimates of project impact are and how they should be interpreted.

A great deal of aggregation has already taken place in producing state level estimates of gains. The threats to the validity of the state level estimates and the pitfalls in inferring that they are the result of Title I are much the same at the state and national levels. The threats to validity are also present with the results at the local level. In addition, the local level results, especially at the level of the individual project, have the uncertainties caused by the limited number of observations.

Small sample sizes lead to large variability in estimates of gain. This is a well established statistical principle and has been well documented in the context of TIERS by Horst (Note 13). For a second grade project with 10 students Horst estimated that a gain (or loss) anywhere between -13 and +13 NCEs could readily be attributed to a variety of factors other than the project, such as school and teacher characteristics, local testing practices, and the vagaries of chance. For larger projects, the band of uncertainty gets narrower, but it is still substantial. The band is also somewhat smaller in higher grades than in the early grades. For example, with 50 participants Horst's estimated band of uncertainty is from -11 to +11 at grade 2 and from -7 to +7 at grade 5.

Even the latter interval of -7 to +7 is large, meaning, for example, that a project of 50 fifth grade students who began the year with an average pretest score at the 25th percentile nationally and ended the year with an average posttest score at the 35th percentile (a gain of 6 NCEs) may not have had a positive effect. Such a gain might easily have resulted from the variety of non-project factors considered by Horst. Furthermore, the majority of the projects have considerably fewer than 50 students. Slightly more than half (51.3 percent) of the 7,608 second grade projects reporting relatively complete data in 1979-80 had their achievement gains estimated on fewer than 10 students. For these projects, even a gain from the 25th percentile to the 45th percentile is not compeling evidence that the project has had a positive effect, since

gains (or losses) of this size are not unusual for groups of 10 or fewer second grade students where no project exists. "The fundamental problem facing local evaluators who wish to measure achievement gains of project students via a norm-referenced model is the large amount of variance inherent in the gains of non-project students" (Horst, Note 13).

Conclusions

1. The predominant pattern of testing project participants in the fall to establish a benchmark against which spring test results are compared to estimate gains is subject to many sources of bias. Although no single source of bias is large encough to invalidate the results, they tend to push the results in the same direction and the cumulative effect is large enough to make the estimated gains overstate the amount of impact by a substantial margin.

2. An annual testing cycle produces results that are less subject to bias than those obtained from a fall-to-spring testing cycle. Even the gains estimated from an annual testing cycle still have a small positive bias, but it is apt to be only about 1 or possibly 2 NCEs. An annual testing cycle would substantially reduce the testing burden for Title I students as well. Title I students on a fall-to-spring testing cycle sometimes are required to take three standardized tests in a single school year: one as part of a regular district-wide testing program (which may be also used in selecting students for Title I projects), a pretest in the fall, and posttest in the spring. The latter two are given solely for the purpose of TIERS. Use of an annual testing cycle could reduce the number of tests from three to one by making use of the existing district-wide testing. The same test could serve as the post-test for the previous year as well as the selection and pretest for the following year. Such use of a single test for all three purposes is made possible by the variation of Model A described in the 1981 User's Guide (Tallmadge & Wood, 1981, pp. 31-35). Use of this approach would not only eliminate the bias inherent in the fall-to-spring test cycle but would even reduce the already relatively small bias in the standard version of Model A with an annual testing cycle.

3. Models B and C and the non-normed variation of Model A suffer severely from either technical inadequacies, practical limitations, or both. Coupled with the rarity with which the variations are used, it is doubtful that the models are worth retaining except, possibly, for reasons of conveying a sense of flexibility and to accommodate a small number of districts.

4. The uncertainties inherent in the estimated gains at the local project level, which typically have fewer than 30 students and about half the time have fewer than 10 students, make it pointless to expect "precise measures of small impacts based on a single fall-to-spring test cycle" (Horst, Note 13). On the basis of data from a single year, they can only be expected to reliably detect exceptionally large effects. The local level estimates of project impact should be accompanied by an indication of the degree of uncertainty of the estimates. This might take the form of stating estimates as a range of values rather than a single number, which gives the misleading appearance of precision. Once a year estimates of performance level are apt to be more trustworthy and more useful than the estimated gains for small districts.

REFERENCE NOTES FOR CHAPTER TWO

1. Stewart, B. L. The Regression Model in Title I Evaluation. Mountain View, Cal.: RMC Research Corporation, 1980.

2. Tallmadge, G. K. An empirical assessment of norm-referenced evaluation methodology. Journal of Educational Measurement, in press.

3. Glass, G. V. Regression effect. Unpublished memorandum, March 8, 1978.

4. Wood, C. T. The adequacy of the equipercentile assumption in the norm-referenced evaluation model. Mountain View, Cal.: RMC Research Corporation, July 1980.

5. Kaskowitz, D. H., & Norwood, C. R. A study of the norm-referenced procedure for evaluation of project effectiveness as applied in the evaluation of Project Information Packages. Menlo Park, Cal.: Stanford Research Institute, January 1977.

6. Roberts, S. J. Differences in NCE gain estimates resulting from the use of local norms versus national norms. Mountain View, Cal.: RMC Research Corporation, September 1980.

7. Elman, A. Quality control in Title I: Manual versus computer conversions of test scores. Palo Alto, Cal.: American Institutes for Research, May 1981.

8. Finley, C. J. What can state education agencies do to improve upon the quality of data collected from local education agencies? Paper presented at the annual meeting of the American Educational Research Association, San Francisco, Cal., April 1979.

9. Roberts, A. O. H. Practice effect of test-wiseness. Mountain View, Cal.: RMC Research Corporation, July 1980.

10. Jaeger, R. M. On the use of standardized achievement tests in Follow Through program evaluation. Greensboro, N.C.: Center for Educational Research and Evaluation, University of North Carolina at Greensboro, January 1981, p. 6.

11. Trismen, D. A., Waller, M. I., & Wilder, G. A descriptive and analytic study of compensatory reading programs. Princeton, N.J.: Educational Testing Service, 1975.

12. Wang, M. M., Bear, M. B., Conklin, J.E., & Hoepfner, R. Student educational development during the school year and the effects of compensatory education. (Technical Report - #10, from the Study of Sustaining Effects of Compensatory Education on Basic Skills.) Santa Monica, Cal.: System Development Corporation, 1980.

13. Horst, D. P. Title I Evaluation and Reporting System: Examination of the models at the project level. Mountain View, Cal.: RMC Research Corporation, March 1981.

CHAPTER THREE

The Technical Assistance Centers

This chapter contains a brief review of the evolution of the Title I Technical Assistance Centers, an assessment of their operation and effectiveness, and a discussion of technical assistance for Chapter I evaluation in the future. Special attention has been given to reporting the magnitude and costs of TAC operations and services; to describing the perceptions of TACs held by the suppliers, receivers, and sponsors of the technical assistance; and to identifying alternative content and modes of technical assistance delivery. This study is a review of the TAC system as a whole. It is not an evaluation of the merits of individual centers.

A supplemental report (Millman, Note 1) has been prepared that contains extensive tables on TAC services and costs; detailed reports of interviews with federal, state, and TAC staff; and a short annotated bibliography of selected literature on technical assistance.

Background

History of the TACs. Section 151 of the 1974 Title I amendments called for OE not only to provide SEAs with models to be used in evaluating Title I programs but also to provide technical assistance to the states to enable them to assist their LEAs in the implementation of the Title I evaluation system. In response, OE established 10 TACs in September 1976. Each TAC provides free technical assistance to all the states within a prescribed region of the country. At the discretion of the states, assistance can also be provided to local districts. With one exception, the same organizations have continued to receive federal awards to operate the TACs through September 1982. Awards in recent years have totaled roughly eight million dollars annually for all 10 TACs combined.

This chapter was prepared by Jason Millman. The assistance of Clinton Walker, James Sanders, James Dunn, Shirley Preston, Cathy Pine, Shukri Sanber, Frank Womer, Todd Rogers, Karen Mitchell, Sharon Davis and David Bridgman is gratefully acknowledged.

The 1978/79 Performance Review of the Technical Assistance Program. In August 1978, OE established an independent panel to evaluate the technical assistance program and to recommend whether the structure and functions of the program should be changed. The principal finding of the report was that the technical assistance program was "working and working well" (Millman, Paisley, Rogers, Sanders, & Womer, 1979, p. 57). Nevertheless, several changes were recommended. Chief among them were:

1. Contracting for technical assistance on a three year cycle;
2. Placing more emphasis on technical assistance directed toward improving programs through evaluation;
3. Increasing TAC staff development;
4. Funding TACs in an equitable manner that includes a consideration of both regional size and regional need;
5. Establishing a uniform system of cost/effort reporting for individual TACs;
6. Coordinating funding for technical assistance on a regional basis (to meet the differing needs of states within a region);
7. Increasing inter-TAC coordination;
8. Coordinating the collection, storage, retrieval, and distribution of TAC-produced materials; and
9. Investigating future systems that have the "flexibility to accommodate to: the diversity of state and local capabilities and needs, and also the enlarged objectives of Title I evaluation technical assistance..." (p. 63)

Methodology of the present study. Three methods of collecting information and opinions were used. First, collegial conversations were held with 35 Title I program and evaluation staff from 29 SEAs, with the director from each of the 10 TACs, with staff from 9 of the centers, and with 10 individuals holding positions in the federal government. 8/ Second, annual, quarterly, and monthly TAC reports, TAC-produced materials, literature on technical assistance, and miscellaneous documents were read. Third, reviews of an earlier draft of this chapter were solicited

---

8/ These conversations were led by Dr. Clinton Walker, Dr. James Sanders, Dr. Frank Womer, and the author.

from members of the panel of the 1978/79 Performance Review, from members of the Planning and Advisory Group of this assessment of TIERS, and from a few other selected individuals. An omission in the design of this study was the failure to seek and interview individuals, especially at SEAs, who might have been less positive than our respondents about compensatory education programs or about evaluation.

Structure, Operations, and Mission of the TACs

The TAC contractors and locations are: American Institutes for Research (Palo Alto, Cal.); Educational Testing Service (Atlanta, Ga.; Evanston, Ill.; and Princeton, N.J.); Powell Associates (Austin, Tex.); Northwest Regional Educational Laboratory (Portland, Ore. - three TACs); NTS Research Corporation (Durham, N.C.); and RMC Research Corporation (Hampton, N.H.). No significant relationships were noted between the type of contractor (e.g., nonprofit vs. for profit) and indicators that were of interest in this study. Differences among individual TACs were, of course, observed; and some of these differences are reported below.

Each TAC is headed by a director and contains, on the average, the equivalent of 11 full time professional staff, or full time equivalents (FTEs). Frequently, one staff member is responsible for the coordination of all services for one state. Most TACs hire staff to serve specialized roles (e.g., a statistician, a media specialist, a computer programmer, a librarian).

TACs serve at the request of the SEAs. A common pattern is for each TAC and its SEA contacts to discuss state needs and possible TAC services and to negotiate the specific assistance that the TAC will provide.

TAC directors gave high marks to the ED project officers and credited them with several changes to the TAC system since the 1978/79 Performance Review. Among these changes are moving to a three-year funding cycle, implementing a monthly management reporting system to track TAC activities and costs, increasing TAC coordination in the development and sharing of printed resources, and establishing a TAC Clearinghouse for storage and dissemination of workshop materials and other documents related to TIERS.

The goals of the TACs are undergoing a large change. At the time of the 1978/79 Performance Review, the TAC objectives were to bring about an

awareness of the requisite evaluation models and their intent, to arrange service agreements, to build confidence and capacity among the clients, and to facilitate implementation of the models. More recently, the TACs have shifted their emphasis to improving the quality of the evaluation data and making these data more usable. The states that are more compliance oriented with respect to Title I evaluation are less likely to include the newer emphases in their descriptions of TAC mission than are states that see Title I evaluation efforts as a way to improve Title I programs. Consequently, it is reasonable to expect that the already substantial variability in SEA use of TAC services will increase once the federal mandate for essentially uniform evaluation procedures is lifted.

Services

Magnitude of the service. Four forms of delivery of technical assistance have continued since the inception of the TACs: workshops, on-site consultations, telephone calls, and written correspondence. In the last two contract years, about 38 percent of TAC staff field visits were workshops; 62 percent were on-site consultations. This is the same ratio reported for the first nine months of 1978 (Millman et al., 1979, p. 21). Four times as many clients were served through workshops than through on-site consultations. As shown in Table 3.1, each professional staff member, on the average, conducted 3 field visits, spent slightly less than 2 days in the field delivering assistance, met 42 clients, and made or received almost 9 substantive telephone calls to or from clients per month. These averages vary markedly among the individual TACs.

Although the total number of TAC contacts is roughly 25 percent greater than reported in the previous review of the TACs, the per person averages seem very low given that the mission of the TACs is to deliver assistance. They also appear low given that both SEA clients and TAC staff emphasized the importance of person-to-person contact.

Allocation of effort. If TAC staff spend only a small portion of their time in the field, what do they do with the rest of their time? Data on time allocations have only recently been required, but it would appear that travel, materials development, telephone calls, correspondence,

Table 3.1

The Magnitude of TAC Field Service

(Nov. 79 - Sept. 80)

	Monthly Rate per TAC		Monthly Rate per Prof. Staff	
	Mean	Range	Mean	Range
Number of Visits				[a]
Workshops	9	4-16	1.1	.5 -3.7
Consultations	15	7-38	1.8	.7 -5.7
All visits	24	14-46	2.9	1.5 -9.4
Staff time in field (days) [b]				
Workshops	6	4-8	.7	.4-1.5
Consultations	10	4-19	1.1	.5-4.0
All visits	16	9-24	1.8	1.1-5.5
Number of Clients Served [c]				
Workshops	281	187-425	33	22-61
Consultations	74	17-178	9	3-23
All visits	355	204-602	42	29-81
Number of substantive tel. calls to and from SEAs/ LEAs	76	38-129	8.9	5.7-13.1

[a] Is read: In the least workshop-oriented TAC, the average professional staff member conducts .5 of a workshop each month; that is, conducts workshops at the rate of one every two months.

[b] Actual time delivering field service. Eight hour days assumed. Travel time is excluded.

[c] The same client seen on different occasions is counted each time.

and staff development together account for one third of their time. Over half of their time seems to be spent on administration and on other activities in support of the technical assistance effort that are not easily identified by the present reporting system.

In terms of number of clients served, March, April, and October are the peak months for field visits. Relatively few visits take place in December and the summer months. Professional staff FTEs, however, remain reasonably constant throughout the year.

The clients being served by TACs differ markedly from state to state. In some states, virtually all the assistance is delivered to the SEA and very little assistance is given to local districts. The reverse is true for other states.

As displayed in Table 3.2, the magnitude of TAC field services, regardless of whether it is to SEAs or LEAs, also differs markedly from state to state. In the 1980-81 contract year, the number of field visits in a state ranged from about 3 to 230 and the number of clients served ranged from about 3 to 3,050. Even when the five most frequently and the five least frequently serviced states were excluded from the analysis, some of the remaining states still received 13 times the visits and had 7 times the number of clients served as other states. When the number of visits and clients served was adjusted for the size of the state's Title I student population, these ratios more than doubled.

Less populous states receive proportionately more field services than more populous states. States rated by TAC directors as being less compliance oriented and trying the most to improve evaluation at the local level receive more field services than states rated otherwise.

What accounts for the substantial variability among states in the amount of field services received? The TAC interviewees suggested that it is the differential wishes of states for technical assistance, rather than any intention of the TACs' part, that lead to the tremendous variability in state use of TAC field services. When the field service data are aggregated across each TAC's member states, however, sizeable differences still remain, which suggest that the variability among states is due <u>in part</u> to the differential penchants of the TACs to

Table 3.2

Distribution of TAC Field Services by States

(Oct. 1980 - Sept. 1981)

Field Visits		Clients Served		Field Visits (per 1,000 Title I Students)		Clients Served (per 1,000 Title I Students)	
No. of Visits	No. of States[a]	No. of Clients	No. of States	No. of Visits	No. of States	No. of Clients	No. of States
180+	3	1800+	5	4.5+	2	45+	5
160-179	0	1600-1799	0	4.0-4.4	4	40-44	2
140-159	2	1400-1599	4	3.5-3.9	1	35-39	0
120-139	1	1200-1399	2	3.0-3.4	1	30-34	2
100-119	3	1000-1199	3	2.5-2.9	2	25-29	2
80-99	2	800-999	4	2.0-2.4	1	20-24	1
60-79	6	600-799	12	1.5-1.9	3	15-19	8
40-59	10	400-599	8	1.0-1.4	9	10-14	6
20-39	14	200-399	9	.5- .9	10	5- 9	18
0-19	10	0-199	4	0 - .4	18	0- 4	7
Total	51[b]	Total	51[b]	Total	51[b]	Total	51[b]
Mean:	60	Mean:	850	Mean:	1.5	Mean:	18
Ratio:	13 to 1[c]	Ratio:	7 to 1[c]	Ratio:	32 to 1[c]	Ratio:	15 to 1[c]

[a] Is read, 180 or more TAC field visits were made to each of three states.

[b] Includes Washington, D.C.

[c] See text.

provide field service. It should be noted in this regard that some states are receiving assistance in ways other than through visits (e.g., through computer programming services). Had such assistance been considered, the reported variability might have been reduced.

Content of the assistance. The TAC message has changed. All respondents agree that significantly less attention is now being paid to how to implement the models. Data for the last two contract years tend to support the claim. Some of the topics that are being emphasized in TAC workshops include improving data quality, managing data, selecting tests, planning evaluations, improving programs through evaluation, and communicating evaluation results to audiences. Workshop materials, some quite elaborate, have been produced to support the shift to program evaluation beyond the TIERS models.

Although all TACs are prepared to offer assistance on any of the topics in the arsenal of TAC materials, the individual TACs have different predispositions toward evaluation procedures. One group, sometimes called the "bean counters," emphasizes data analysis and technical matters of evaluation. The second group, sometimes called the "fuzzies," emphasizes the less quantifiable aspects of evaluation.

Restrictions. Federal limitations preclude the TACs from conducting evaluations, recommending specific tests, advising on curriculum and programs, and enforcing or advising on government policy. We found widespread agreement that the original restrictions were wise, for they protected the TACs and gave the program a consistent focus. Restrictions are seen as a powerful and accepted way for ED to keep control of TAC services.

Assigned tasks. ED provides financial support to individual TACs to perform commissioned tasks that benefit the technical assistance effort as a whole. These assigned tasks include operating a Test Information Center, developing workshop materials on designated topics, and maintaining a Clearinghouse of Title I evaluation training materials and assorted related documents.

The Clearinghouse was established in the spring of 1979 at the TAC for Region VII (Palo Alto, Cal.) to promote sharing of these Title I

related materials and documents among the TACs and to reduce duplication of TAC efforts. The collection has grown to over 1,200 documents; that size is limited by the lack of money to process the material rather than by a lack of suitable items. Listings of catalog descriptions are periodically sent to over 130 potential users.

The TAC Clearinghouse staff see themselves as librarians. The staff are a bit frustrated because some TACs are not promoting the Clearinghouse service as much as they could and because the Clearinghouse staff themselves cannot work directly with clients from other than the TAC Region VII member states.

Most directors of the other TACs state that the Clearinghouse produces work of a high quality but feel it is a more elaborate and costly resource than is needed.

Approximately one-sixth of the TAC Region VII budget is spent on the Clearinghouse. With only 240 requests last year, the unit cost of service is extremely high. Most of the variable cost is in entering documents into the system; the variable cost of fulfilling a request is very low.

Many of the TAC-produced materials have worth beyond Title I and deserve to be saved in a public archive like the Clearinghouse. However, it does seem that a return to the more modest functions originally intended would be financially prudent.

Costs and Effectiveness of the TACs

Of every dollar expended, about 45 cents is for salaries, 14 cents is for travel and other direct costs, and 41 cents is for indirect costs and fees. Expenditures by individual TACs are not proportional to the Title I student population in the respective regions nor to the TAC's amount of field service. The TACs' total expenditure reflects, as well, such contributions to the support of the overall TAC system as materials development, the Clearinghouse, and the Test Information Center. More recent allocations to the individual TACs appear to have had a more rational basis than the allocation made in 1977/78.

Because of the cost and effort reporting requirements initiated after the 1978/79 Performance Review, it is possible to estimate the unit cost

of TAC field services. As shown in Table 3.3, average costs are roughly $2,400 for a field visit, $480 per hour for instruction, $170 per client served, and $44 per client consultation or training hour. Rates for individual TACs vary greatly.

In calculating the dollar figures in Table 3.3, services other than field services (e.g., development of computer programs or training materials used by the clients) were excluded. Thus, in this sense the dollar amounts overestimate the cost of total client services. In another sense, however, the dollar amounts underestimate the total costs, since the field services require additional expenditures by ED and by the clients themselves that were not considered in the calculations.

One of the limitations of this study is that it lacks extensive comparative cost data that can be used to assess whether the costs reported are high or low. One recent evaluation (Walsh, Note 2) of a decentralized technical assistance program in the medical field, however, reports costs of $1,277 per visit, $39 per client, and $4.15 per client service hour. The program is not strictly comparable to the TAC program, since all services were in a group setting (mean size was 34 participants); no individual consultations were provided.

"Consultants" working out of their homes or university faculty working off campus certainly do not ordinarily charge $480 per hour (or $3,840 per day) for their time. Nor do SEAs or LEAs customarily compensate consultants at that rate. Similarly, it would be interesting to speculate whether the typical client group would be willing to spend $2,400 for a consultant visit.

In judging the costs for TAC field services, qualitative considerations deserve to be kept in mind. The TACs' job has not been easy. Their SEA and LEA clients often were initially hesitant, inexperienced, and overburdened. In the beginning, this state and local audience saw little hope of reward for accepting help from the TACs but rather envisioned the prospect of an absence of penalty. TAC clients are geographically dispersed and not able to travel freely. Turnover of the client staff is high and continuous. The TACs' goal is to change the practice of clients, which requires repeated person-to-person contact, individualized services,

Table 3.3

Cost of Client Services

(Nov. 79 - Sep. 81)

TAC	Each Field Visit	Each Hour of Providing Assistance in the Field	Each Client Served in the Field	Each Client Hour of Service in the Field (a)
---	$1090	$230	$130	$36
---	1270	310	180	58
---	2000	320	90	32
---	2660	470	180	56
---	2680	450	110	41
---	2920	570	190	51
---	2950	560	200	49
---	2950	850	250	85
---	4900	770	250	91
---	5180	900	190	75
All TACs (b)	$2430	$480	$170	$55

a Estimated using a 3-predictor regression model to extrapolate from recently acquired client service hour data to the November 79 - September 81 cost and service data base.

b Weighted average.

and a back-up for the TAC staff of a range of human and material resources. The cost of TACs must be evaluated in the context of these obstacles and these requirements.

The typical SEA Title I evaluator is unquestionably satisfied with the TAC performance. The TACs and their work receive outstanding ratings from SEA personnel. TAC staff are viewed as helpful, competent, likable, objective, responsive, and professional. The high regard and enthusiasm for TAC personnel and services implied in the 1978/79 Performance Review show no signs of letting up.

A lengthy list of TAC accomplishments was suggested by the interviewees. In addition to facilitating the implementation of TIERS, several respondents mentioned as positive results attributed to TAC efforts the following: raising of client consciousness about evaluation, facilitating interactions among evaluators from different states, improving the clients' capabilities as evaluators, being responsible for the improved quality of information provided by test publishers, and developing evaluation materials valuable for people other than Title I evaluators.

The interpersonal skills of the TAC staff are seen by the SEA respondents to be, by far, the most important factor that facilitated the effectiveness of TAC services. Descriptions of this factor take many forms, such as good personality, willingness to learn about the special needs and differences among constituencies within the state, and becoming a known and trusted partner. The availability of technical assistance almost any time it was needed is applauded by several respondents. Physical proximity is <u>not</u> widely viewed as important for TAC effectiveness.

Some SEA respondents feel that the TACs can be most effective with larger states and districts, or with those having more staff who care about evaluation, since they would be more receptive to technical assistance and better able to benefit from it. Other SEA respondents present a different view. For them it is the small states and nonurban LEAs who, because they do not have staff sophisticated about evaluation, have the most to benefit from technical assistance.

Technical Assistance in the Future

Opinions of the respondents. As a group, the SEA respondents have a positive attitude toward evaluation, have a high regard for its worth, and see a continuing need for assistance in conducting evaluations. This investigator also shares these views, and the remainder of this chapter never questions whether there should be technical assistance for Chapter I evaluation but only questions what its content might be and how it might be delivered.

As noted above, TAC work is shifting from an emphasis on TIERS to other, more general evaluation activities. How far should the content of the the technical assistance expand? Most respondents believe that the TACs' assistance should be limited to evaluation and not be expanded to include providing advice on specific curricula or instructional practices. Giving advice on evaluation and giving advice on curriculum and instruction are separated by a fine line. Many TAC staff feel confident that they can provide evaluation assistance close to the line by offering a rational approach to curriculum planning, curriculum selection, and curriculum implementation and management. TAC directors tend to be somewhat more cautious than many of their staff or many state evaluators, and they argue for moderation. For many of them, effectiveness requires focus. To them and to this investigator, it seems prudent that TAC efforts for the immediate present be confined to those aspects of evaluation that, by virtue of training and experience, the TAC staff can do best.

It is less clear to the respondents how far assistance in evaluation should venture beyond Chapter I programs. In most LEAs, only a few students per grade are served by Title I 9/, which suggests that if assistance is to be efficient for these districts, a targeted student body broader than Chapter I is required. In contrast, in every state a few large LEAs account for a high proportion of the targeted population, which suggests that if assistance is to be efficient for these states, a targeted student body of Chapter I students from those LEAs is required.

9/ For example, in about two-thirds of the projects in Montana, Title I serves only five or fewer students per grade.

The differing needs and interests for assistance at the state level have been mentioned as have the individual TAC differences in inclination for type and amount of field service. Balancing this diversity of numbers, needs, interests, and inclinations against the desire for a focused program of optimum quality may require new guidelines and perhaps new structures for the delivery of assistance.

The state respondents like the TAC system the way it is. They like the individual attention and follow up service they receive, and they appreciate the technical resources to which individual TAC staff members have access. When presented with other systems, the state respondents were very negative. A single, national TAC was seen as too impersonal; a high technology system as too infeasible; a TAC person housed in the SEA as too vulnerable to compromising duties; and a system where the SEA pays the servicer of choice as too impractical because the states would not spend the money. 10/

Features of the current TACs that may no longer be essential. The history of the TACs shows many attempts to refine the system. The TAC structure, however, has remained virtually untouched. Six years have passed since the TACs were established and new conditions (e.g., no federally mandated Title I evaluations, clientele more aware of the TACs, new attitudes in Washington, etc.) may now make certain features (or assumed features) of the early TACs no longer necessary. In planning a technical assistance system for the future, it is important not to assume restrictions on design that may not, in fact, be necessary or desirable. To say that TAC features may no longer be essential, however, is not to say that it would necessarily be wise to change them.

In each of the next paragraphs, a feature (or assumed feature) of the TACs is stated first, then commented on.

1. TAC fiscal support is a fixed amount during the contract period, regardless of the demand for, or the amount of services delivered. This fixed funding was reasonable when early start-up and materials development costs were high and when promoting the need for evaluation services

---

10/ For another opinion about the delivery of Title I evaluation assistance, see Apling and Kennedy (Note 3).

had to be supported. Responding to requests, rather than promoting services, is now the professed present modal pattern.

2. TAC services are provided free of charge. Title I is an entitlement program. It was argued that SEAs and LEAs would not perform a complex, esoteric, mandated evaluation if they had to pay for learning how to do it. Chapter I no longer requires elaborate or uniform evaluations.

3. TACs have protected, regional territories. TACs were given exclusive territorial rights. Young enterprises often need special advantages that mature ones do not.

4. TAC work is provided to LEAs only with the approval of its SEA. The mandated communication link between federal, state, and local agencies for reporting Title I achievement results based on a common set of evaluation models has been weakened by Chapters I and III.

5. TACs provide services equally to all states. When TIERS was mandated, it was expected that states would receive proportional amounts of service, depending on need, size, and other explicitly stated criteria. Requests for and delivery of service among client groups are now very uneven. When Chapter I evaluation becomes a state and local option, the variation in requests can be expected to increase even more. A system that responds to client requests, rather than promotes its services in an attempt to satisfy some implied formula for distribution of effort across states, would seem more in keeping with today's Washington philosophy.

6. TACs differ only in their geography, not in the services they provide. It is already the case that the services vary across TACs, and this variability can be expected to increase even more. With the models gone, a single story no longer needs to be told; standard procedures no longer have to be implemented.

7. TACs are workshop/training oriented. Workshops, and materials developed in support of them, are most efficient and appropriate when there is a fixed body of knowledge or techniques (like how to implement the models) to be disseminated to a large number of participants. Needs for technical assistance in the future are more likely to be unique to the specific situation and limited to fewer clients.

Alternative systems. The features of the TACs could change radically in the future. As an example of just how very different a technical

assistance system might look once the features (or assumed features) listed above are relaxed, consider the following. Technical assistance could be provided via a national "brokerage agent" and a set of eligible "bidders" (perhaps a dozen or so "approved" organizations). SEA and LEA clients would submit requests for services to the brokerage agent, who in turn would immediately circulate those requests to the eligible bidders. For modest service requests, interested bidders would respond merely by letter or prospectus directly to the client. The provider would be chosen by the client. For larger service requests, the bidder would respond by proposal and awards would be made with the help of an external review. The federal government would pay most of the cost of the service; the SEA or LEA client would pay only a small portion of it. The brokerage agent would be responsible only for distributing the work requests promptly, for public disclosure of all successful bidders, and for fiscal detail.

This hypothetical system is intended to illustrate what kind of a technical delivery scheme might result from the simultaneous abolition of all seven TAC features (or assumed features) listed above. Less radical departures from the current TAC system abound. As one example, the protected territory and equal service features could be relaxed by permitting each provider to offer its own specialized services that could be contracted for by clients outside its region. As a second example, the free service and fixed TAC support features could be relaxed by requiring clients to subscribe to their TACs for heavily subsidized services in advance of ED contract awards, which would be based on the amount of requested service.

It is tempting to explicate such hypothetical systems more fully. Further discussion of them is avoided, however, because the reason for mentioning such systems is <u>not</u> to recommend them, but rather to illustrate how the features of the TAC system of the future could be very different from the features of the present TAC system.

<u>TAC Characteristics (Plus and Minus)</u>

In any discussion of how a system could be improved, it is essential to keep in mind the characteristics of the system, as it currently exists.

that enhance as well as detract from its effectiveness. On the positive side, the TACs as they are currently established:

1. Have an established network of professional contacts that in many instances span a number of years;
2. Have a record of well received service;
3. Have a cadre of trained staff;
4. Are well liked by the majority of their constituents, who are likely to voice some concern if the system is changed;
5. Are a known entity. (There is less risk in keeping the status quo than in switching to another system.)

On the negative side:

1. The cost of technical assistance has been high for the magnitude of services provided. (The TACs are fully subsidized operations, and the clients have little investment in the assistance provided.)

Concluding Thoughts

Four clear and major findings emerged. First, the content of the assistance has shifted from implementing the TIERS models to improving the quality and utility of evaluation data. Providing such assistance, which is anything but straightforward, places high intellectual demands on the TAC staff and highlights the importance of the back-up personnel and materials resources of the system.

Second, the magnitude of services received varies widely from one state to another; the amount of field service offered and, to a lesser extent, the type of assistance provided, vary widely from one TAC to another. Captializing on this diversity rather than eliminating it is a desirable option. Uniformity is not an end in itself.

Third, the amount of field service is low relative to the money expended. Changing the delivery of technical assistance in ways that would decrease the unit cost of the services is warranted, provided reaching quotas does not replace achieving quality service as a TAC goal.

Fourth, the TAC clients are well satisfied with the TAC services and want them to continue. A cynic could ask, "Why not? The services are free. Are the clients sufficiently enamored with the assistance

that they would pay five or ten cents on the dollar for it?" The cynic has a point, but no service comes free. As one client put it,

> These TAC people are quite good, thank the Lord, or we could not be using them. We don't need the aggravation. If they were not like that, they would be costing us more than they were worth.

The TACs could have blown it and been turned away. Rather, they have done what was asked of them and done it in a way that has most clients wanting more. That achievement is worth being proud about.

REFERENCE NOTES FOR CHAPTER THREE

1. Millman, J. A descriptive account of services, costs and opinions about the Technical Assistance Centers for Title I evaluation. Palo Alto, Cal.: American Institutes for Research, Title I Evaluation Clearinghouse, 1982.

2. Walsh, P. L. Interwest Regional Medical Education Center FY 81 progress report. Salt Lake City, Utah: Veterans Administration, Interwest Regional Medical Education Center, circa 1981.

3. Apling, R. N., & Kennedy, M. M. Providing school districts with assistance to improve the usefulness of tests and evaluations. Cambridge, Mass: The Huron Institute, September 1981, pp. 82.

CHAPTER FOUR

The Utility of Title I Evaluation

There is increasing concern with the degree to which program evaluations have practical utility for educational decisionmakers.* Utility in the broad sense refers not only to actual utilization but also to the potential for utilization. In the present chapter, we examine instances in which utilization has occurred at the SEA and LEA levels, and we examine the contextual factors that have contributed to this utilization. Our intent is not to measure systematically and exhaustively the level of utilization. Rather, by identifying instances of utilization, we attempt to determine whether the Title I evaluation system has utility.

We found that all of the SEAs use evaluation data to monitor LEA compliance with Title I regulations. Beyond that, most also use the data to help LEAs identify and resolve problems. And not infrequently, Title I evaluations have a "spill-over" effect into other areas of state level educational planning and evaluation. The LEAs we studied demonstrate a wide range of utilization by district school boards and advisory committees, individual school administrations, local parent committees, and classroom teachers. The uses fall into four broad categories: organizational change, instructional action, problem recognition, and attitude change.

We used a broad definition of utilization in this study. The review of research that is summarized in this chapter suggests that in studies of Title I evaluation use, as in other utilization studies, restricting the definition of utilization to identifiable impact on specific actions ignores a great many important uses for evaluations. The perspective on evaluation use this research team subscribes to was described by the author in an earlier work (Alkin, Daillak, & White,

---

*The authors of this chapter are Marvin C. Alkin, Brian Stecher, and Frederica Geiger. Larry Braskamp, Richard Daillak, Jean King, and Richard Williams assisted in conducting the field studies.

1979). Utilization may include gradual incremental influences on attitudes and judgments throughout the entire course of decision formulation. This conception may be contrasted with the narrower view which considers utilization in clear-cut terms at the actual point of decision. A review of the relevant research illustrates how this perspective has informed the studies which preceded and shaped our own.

Previous Research

Only a small number of studies, published between 1977 and 1981, have examined Title I evaluation specifically. They vary considerably in focus, strategy, intent, and depth. The following studies, listed chronologically, were examined in our literature review:

- The Syracuse Research Corporation (Goettel, Kaplan, Orland, Forgione, & Huff, Note 1) and
  Follow up Study (Forgione, Kaplan, & Orland, Note 2);
- SRI International (David, Note 3);
- Hamilton/Chicago (Note 4);
- Northwestern University (Boruch & Cordray, Note 5);
- The Huron Institute (Kennedy, Naumann, & Apling, Note 6);
- The State of Indiana (Crowley, Note 7);
- The RMC Corporation (Cichon & Herr, Note 8); and
- Center for the Study of Evaluation (Stecher, Alkin, & Flesher, Note 9).

The pattern of results in the various studies on the use of Title I evaluation information is clear and quite consistent. They confirm the finding that Title I evaluation is primarily directed toward reporting and compliance functions. However, early appraisals that this was the only use of Title I evaluation have been moderated. More recent research offers an expanded point of view. Though TIERS remains a compliance oriented system, the literature provides evidence that local use is being made of other Title I funded evaluation data. The research indicates that various types of supplemental Title I evaluation help local school personnel recognize areas in need of attention and confirm impressions obtained through other sources. In addition, there is evidence of other indirect and incremental uses of Title I evaluation (of all types) which

are more difficult to characterize (Cichon & Herr, Note 8; Stecher et al., Note 9). Overall, these studies indicate considerable use of supplemental Title I evaluation information for problem recognition, through there is less evidence of more direct use of evaluation data.

The recommendations for increasing evaluation use made in these studies, or implied in them, have particular bearing on this study of utility. Forgione et al. (Note 2) believed that an increased awareness of evaluation as "a serious and complex responsibility" is a necessary precursor to changes in utilization. Kennedy et al. (Note 6) and Crowley (Note 7) reiterated the point. Boruch and Cordray (Note 5) supported the contention of Alkin (Note 10) that utilization can be enhanced by encouraging districts to supplement the minimum TIERS requirements with evaluation activities directed to local needs. Hamilton (Note 4) stressed the importance of communication between program personnel and evaluation personnel. Boruch and Cordray (Note 5) and Kennedy et al. (Note 6) made the point that the needs of local personnel are important determinants of use. Many of the preceding suggestions for increasing evaluation use were supported in this study as well and will be examined in a later section.

A more fundamental issue, however, has been posed by David (Note 3), who suggested the following strategy for enlisting technical assistance to improve utilization (pp. vii and ix):

> The USOE evaluation models are designed to improve the quality of the data and will not, by themselves, lead to an increase in the local use of evaluation. However, the current technical assistance strategy, if redirected, could serve as a powerful force in changing how evaluation is perceived and thereby increase evaluation use locally. To accomplish this goal, technical assistance must be redesigned to communicate a new view of the role of evaluation and to develop skills such as generating one's own evaluation questions. As long as technical assistance is defined narrowly as a way of telling local staff how to improve the quality of their data, it will not increase local use of evaluation.

The suggestion that expanded technical assistance offers the greatest hope for increasing the influence of TIERS data on local program decisions

has, in fact, been implemented. As Stonehill and Groves (Note 11) report, since 1979-80 efforts of the Title I TACs have focused on local utility of TIERS data. Earlier TAC efforts were only indirectly aimed at increased evaluation utilization. These efforts concentrated on producing the high quality data which it was felt could have credibility in the eyes of local audiences.

One would anticipate that the recent efforts of TACs to help LEAs could have a positive impact on the use of Title I evaluation for program improvement. Thus, the current study might provide insights on TIERS use not afforded by prior research.

Methodology

Our strategy was to study SEAs and LEAs with exemplary Title I evaluation utilization. Therefore, the project director began by interviewing the directors of the 10 regional TACs to discuss with them the nature of evaluation and evaluation utility in their regions and to learn from them what states might be fruitful for further analysis. The extensive telephone interviews also resulted in a list of SEAs recommended for their high utilization. From these, 12 SEAs (three in each of four geographical divisions) were selected for interviewing.

The data collection could, due to time constraints, last only five weeks. At the SEA level we interviewed the Title I contact person by telephone. We used open ended interviewing procedures, which could be most responsive to eliciting subtle evaluation influences. The discussions with SEA personnel all reflected a similar agenda, but specific questions and requests for subsequent clarifications varied from interview to interview depending upon the particular comments of the SEA respondent. The following topics represent the kinds of concerns addressed in each interview. They also reflect the general interview agenda employed:

1. Introduction: indicating who we were, what we were doing, and why.
2. Overview of SEA evaluation and utilization: encouraging the respondent to discuss the nature of Title I evaluation in

his or her SEA, including general thoughts on instances of utilization.

3. Problem recognition: encouraging the respondent to consider and discuss situations in which evaluation information might have contributed to the recognition of a problem but not necessarily to its solution.
4. Problem solution--administration: expressing an interest in the extent to which statewide Title I evaluation data were a source of information related to modifications in administrative procedures at the state level.
5. Problem solution--instructional improvement: expressing an interest in whether instructional changes or other substantive changes in the Title I program might have been affected by statewide Title I evaluation data.
6. Factors affecting utilization: exploring contextual, political, or personal factors that seemed to affect statewide evaluation utilization.

Typically, one or two individuals were interviewed within each SEA, beginning with the official designated as the Title I contact person and sometimes progressing to other personnel. In many interviews, the respondent elected to collect materials, check his or her files, and discuss issues with colleagues in order to clarify certain points that had been raised in the interview.

Both TACs and SEAs were asked to recommend LEAs with outstanding forms of Title I evaluation use, and we selected 20 LEAs from this list. Four researchers who had published in the utilization field and who shared our broad perspective on utilization were selected to conduct interviews. A two stage interviewing procedure was adopted. In the first stage, five interviews of this type were conducted by each of the regional researchers, and the information gleaned from these was then used to identify two LEAs in each region for further intensive telephone interviews and one additional LEA in each region for a district site-visit. Thus, in the second phase of the LEA interview process, three LEAs in each of the four regions were selected for closer scrutiny.

The subsequent in-depth interviews lasted approximately an hour per interview. Respondents were selected from seven categories: (1) Title I coordinator; (2) Title I evaluator; (3) Superintendent or Assistant Superintendent (curriculum, instruction, program, etc.); (4) school board member; (5) principal at one selected school; (6) Title I coordinator or other individual primarily responsible for the Title I program at the same school; and (7) parent member of the school advisory council. Of course, in smaller LEAs, some of these categories did not exist, or a single individual acted in more than one role. The site visits lasted a day and a half and focused on a similar, but larger cross section of people. In some cases, additional personnel were contacted by phone after the site visit.

The phone interviews and site visits were recorded on summary data forms prepared for this study. In addition, the site visit notes were developed into case study narratives. Attention in these narratives was devoted particularly to descriptions of evaluation use and to contextual factors that seemed relevant to high levels of utilization.

Evaluation Activities and Uses — SEA Level

Activities. We found three major kinds of evaluation activities at the SEA level. Working with an expansive view, we found that evaluation is not simply restricted to the collection and analysis of TIERS data. Functions also include various monitoring activities conducted by the Title I staff and technical assistance to LEAs.

We learned that collecting TIERS data is viewed as the primary responsibility of SEA evaluation personnel. However, how that responsibility is discharged varies widely among the 12 states. In four states evaluation belongs to a non-Title I evaluation office, in the other eight to the Title I office. In four cases, one individual has specific evaluation responsibilities; in other cases evaluation teams with other responsibilities conduct TIERS evaluation, and in still others, outside experts or the program director takes responsibility.

Furthermore, SEAs collect different kinds of data. About one third of the SEAs go beyond the basic requirements of TIERS. Some analyze the data in great detail, breaking down cost data by several subcategories

or collecting data on an individual student basis rather than by school or district, for example. Several states add entire categories of information to their surveys, such as the extent to which a school has a management plan and an internal assessment program for monitoring student progress.

The monitoring functions are typically performed by consultants within the Title I office -- although there are a number of minor variations on this theme. In several instances, monitors have regional responsibilities. In other cases, the staffs available for monitoring are quite small and hold other responsibilities as well. Monitors visit schools primarily to insure program compliance; however, monitoring activities are also designed to provide program advice and consultation. Typically, the specific advice provided emanates from the expertise and experience of the monitors rather than from the evaluative data.

A variety of evaluation-related technical assistance activities are provided by SEAs, in all instances organizationally located within a program oriented office. One of the most prevalent types is designed to insure that the TIERS data are collected and provided properly by the LEAs. Other technical assistance is related to developing the district applications, particularly the evaluation-related portions of the application (such as the needs assessment data). Technical assistance is provided for compliance issues such as the assurance that student selection at the LEA level is taking place in an appropriate manner.

We found one other type of technical assistance to evaluation. Some SEAs have engaged in development of procedures, instruments, and forms which might be of value for local information purposes. For example, one Southern state interviewed during the course of this study has devised a set of 22 data collection charts that districts might use in conducting process evaluations.

Use. We have distinguished three categories for state level uses of evaluation information: compliance, situation recognition, and administrative/curricular actions.

All states studied monitor LEA compliance with Title I regulations. Thus, the accuracy of achievement data receives the greatest attention.

The SEAs have devised a variety of methods to improve the quality of achievement data; some directly reviewed data collection procedures. Almost all SEAs are also concerned with proper use of the three Title I models and with student selection.

Most states use TIERS data in conjunction with other supplemental Title I evaluation data to recognize situations demanding attention -- whether they are "problems" or exemplary instances to be rewarded or showcased. A common method is to distribute summaries with comparative district-by-district data. Sometimes the SEA role ends when the data are communicated to LEAs, but more often SEAs do more. They might, for example, "flag" particularly low scores and call them to the attention of regional program offices in order to prompt LEA action. The same comparative data also spotlight exemplary programs, of course, and at least one Eastern state uses the data to encourage the "successful" programs to apply for recognition by the Joint Dissimination Review Panel (JDRP) -- and then assists them in applying.

More than half of the SEAs have used Title I evaluation as a factor in taking an action, generally in one of three ways:

1. Initiating programs of inservice and technical assistance to LEAs;
2. Making programmatic recommendations to districts based on systematic study of factors that affected educational outcomes;
3. Using those data as input to some other educational question that was being decided at the state level.

An example of technical assistance uses occurred in a Western state where the evaluator sat in on planning sessions and brought knowledge of the aggregated Title I performance data to bear on relevant issues. A number of other states use Title I evaluation to locate areas of need for technical assistance. After a problem has been recognized, however, SEAs typically rely on other information sources; thus, advice about the solution more often follows from on-site observation.

SEAs use Title I information in many ways to analyze program characteristics and make recommendations, generally in conjunction with

other evaluation data. Some states base their recommendations on casual study of the data; others conduct systematic review and formal analysis before determining their preferred course of action. Some states offer these guidelines as informal suggestions to LEAs and allow each district to decide for itself whether it will adopt the particular recommendation; other states act in a more persuasive fashion.

Finally, SEAs frequently reported that they use trends in Title I evaluation data as a basis for making educational decisions about children not specifically served by Title I projects. Similarly, Title I evaluation procedures have been imitated by SEAs in other evaluation systems.

A number of factors appeared to influence the extent of evaluation use at the SEA level. Administrative responsiveness to evaluation information was clearly one of the most important. A close relationship between the evaluator(s) and administrators -- usually fostered by placement of the evaluators within the program unit -- was another. It was also clear from our interviews that TACs play an important role in shaping attitudes toward evaluation. By educating both SEAs and LEAs about the quality of TIERS data and helping to improve testing procedures, TACs have fostered a willingness to use the data.

Evaluation Activities and Uses -- LEA Level

Both the literature review and the LEA interviews suggest similar conclusions about Title I evaluation utility at the local level. While the reporting and compliance functions of Title I evaluation predominated in most earlier studies, some recent evidence indicates other forms of data use at the local level -- especially since TACs have placed increased emphasis on utilization. In our interviews, LEA personnel reported using Title I evaluation data to change attitudes and opinions, to recognize situations requiring attention, and to contribute to decisions on organization and instruction. We did not uncover dramatic instances in which evaluation data acted as the major force in a significant program decision. Instead, we found a multiplicity of situations in which Title I evaluation played a part in small actions and in gradual, incremental program evolution on a larger scale.

Table 4.1 presents the instances of evaluation use mentioned in the field interviews, organized by type of utilization and type of evaluation information. In brackets, we have recorded the duplicated instances, Thus, "4[7]" refers to four different events, some of which were mentioned more than once for a total of seven citations.

Typically, the local uses of Title I Performance Reports and TIERS for attitude change involved attempts to persuade others about the value of the Title I program. Examples of this phenomenon include the following:

- A Title I Director in an Eastern LEA, who used summary TIERS statistics in his presentation to teachers to help them formulate more reasonable expectations for student performance.
- A Title I Coordinator in another Eastern district who used TIERS pre-post achievement data to demonstrate the success of the Title I program to members of the school board. She hoped that increased local funding could be obtained to offset anticipated declines in federal support.

In most instances the local use of Title I Performance Reports and TIERS for situation recognition involved the discovery by the Title I Coordinator or Title I Director of previously unknown facts about the program. Examples of this situation include the following:

- A Title I Coordinator in a Southern district who discovered a surprisingly high rate of absenteeism when he reviewed the Title I Performance Report. He communicated this information to the Title I director, who had previously not recognized the severity of the problem.
- Teachers and parents in a Southern district who jointly reviewed students' Title I performance scores to better diagnose problems so they might consider ways for parents to assist with their child's education.

The following examples are typical of the kinds of local administrative and organizational changes that are informed by TIERS evaluation data:

- A Title I Coordinator in a Midwestern district who was prompted

Table 4.1

Instances of Evaluation Use in Exemplary Districts

Number of Incidents [duplicated]

Type of Utilization	Type of Evaluation Information: Needs Assessment	TIERS	Local Process Evaluation	Testing For Student Selection	Other LEA Testing	Other	Total
Attitude Change	4 [4]	15 [18]	8 [10]	1 [1]	9 [12]	4 [4]	41 [49]
Situation Recognition	2 [2]	24 [28]	5 [7]	2 [2]	3 [4]	7 [8]	43 [51]
Administrative or Organizational Action	4 [4]	18 [20]	10 [11]	6 [8]	6 [7]	4 [4]	48 [54]
Curriculum or Instructional Action	2 [2]	20 [29]	11 [11]	3 [3]	7 [10]	3 [5]	46 [60]
Total	12 [12]	77 [95]	34 [39]	12 [14]	25 [33]	18 [21]	178 [214]

by TIERS reading scores to gather observational data on the operation of both the Title I reading program and the regular reading program. He then took administrative actions to improve articulation between the two programs.

- An assistant superintendent in a Western district who used the data from TIERS to make administrative decisions on program cuts. As he put it, "When you have an expanding program you can get by with limited evaluation data, but when you have to make cuts the data become very valuable."

The following examples typify instances in which local curricular and instructional decisions were based on TIERS data:

- A Title I Reading Coordinator in a Midwestern district who, based on unusual patterns in the test data, changed the criteria for grouping students in the reading program and selected new materials.
- A principal in a Southern LEA who based his decision to change a major aspect of the school's reading program on the declining test scores he observed in the TIERS reports.

In all, over 100 instances of LEA evaluation use (in 12 LEAs) were described to us. They illustrated the fact that evaluation data have utility at many different levels in an LEA. More important, we found that different patterns of use predominate at different organizational levels in the school district. Numerous examples are cited in Alkin, Stecher, and Geiger (Note 12).

To some extent there were also differences from level to level in the balance between the use of TIERS data and the use of other Title I evaluation data. Summative data, such as TIERS, were predominant sources of information in large policy decisions made by school boards, district advisory committees, and other external agencies. In many of these cases the administration had used data to persuade the advisory groups to behave in a certain manner.

At the district administrative level, on the other hand, TIERS data were mixed almost equally with other Title I evaluation information developed within the district. At the school building level the mixture of data types was similar, though there was somewhat more reference to

TIERS. Finally, at the classroom level a wide variety of tests and evaluation information was useful. The more closely geared the information was to the instructional program, the more likely the classroom teachers were to make use of it. TIERS data were of less utility to classroom teachers than other evaluation data more closely linked to the instructional program.

Factors Leading to Utilization: The Case Studies

The literature on evaluation use (both general references and research specifically focused on Title I) and our interviews at SEAs and LEAs offer a number of suggestions about what factors increase utilization -- that is, what gives evaluations utility. We have identified six major clusters of factors which will be illustrated with examples from the LEA case studies developed by the four LEA interviewers. (The entire case study reports are appended to Alkin et al., Note 12).

The following chart presents relevant details on the case studies. It should be noted that the pseudonyms we gave each study reflect the geographic regions and demographic size of each site.

Table 4.2

Case Study Information

CASE STUDY	INVESTIGATOR	EVALUATOR(S)	TITLE I GRADE LEVEL
Middle City, Midwest	Larry Braskamp Univ. of Illinois	Jack Brown	K-12
Big City, East	Richard Daillak St. John's Univ.	Harold Green	1-8
Small Town, South	Jean King Tulane Univ.	Sally Lopenia Bess Foley	K-8
Suburb, West	Richard Williams UCLA	Bill Leonard	K-12

Evaluator credibility. The credibility of the evaluator was consistently cited in the case studies. Forcefulness and earned respect were the qualities that successful evaluators possessed. King noted that the effective use of Title I evaluation information in Small Town was "a tribute to the skill and hard work to two women [the evaluators]." One, "as everyone around her is quick to point out, [is] a positive force in the local school community." About the other woman, King noted, "One teacher labeled her a 'unifying force' and the 'only source of information for our Title I program.' The parents and principals also shared this feeling, suggesting the potent effect of a single person's commitment to the use of information in a school district."

Williams, talking about Suburb, West's evaluation use, described "the evaluation specialist everyone credits with developing Suburb West's evaluation system." Daillak found that in Big City, East, "Mr. Green undoubtedly helped persuade the region's administrators of the value of evaluation" and continued, "Mr. Green's university training prepared him for the technical demands of his job, but his personal characteristics truly suited him to it."

According to Braskamp, the evaluation director in Middle City, Midwest, was respected in the district even though his information sometimes caused resentment. The assistant superintendent told Braskamp that the evaluator is "brutally honest. He wants the truth to be known."

Evaluator interest in utilization. The attitudes of the evaluator toward evaluation use have direct bearing on utilization. The evaluators in our case studies had strong beliefs in the importance of evaluation utilization. In Suburb, West, for example, Williams found, "Mr. Leonard has a fundamental belief that evaluations are a waste of time unless they are used by someone." Small Town's evaluators began with a slightly different slant, "but (their) attitude was a positive one, i.e., given that we have to do these things, how can we make them work for us?" As a result of this positive attitude, the evaluators "sought to make the experience worthwhile and valuable." King asserted that the result

of this early positive attitude was that, "Both Mrs. Lopenia and Mrs. Foley want people to use the information they gather; the test scores are 'data worthy of being used and shared.'"

In Suburb, West, Leonard worked to create positive attitudes toward evaluation and thereby increase its utilization. The effectiveness of his approach is illustrated by one of the changes in Suburb, West, which Williams describes. "In the past, school personnel feared data, believing that somehow it would hurt the school or be used against it. Slowly the schools have learned that data can be used to identify and solve problems." Similarly, Daillak noted that in Big City, Mr. Green met with union stewards for all of the schools. "At that meeting, he explained his 'ground rules' for evaluations -- anonymity for individual teachers and schools, and independence from building principals." By taking this approach, Green was able to "secure the cooperation of the union." Teacher attitude toward evaluation became more positive and less fearful.

Interest of decisionmakers and the community in evaluation. The interest and attitudes of decisionmakers and the local community toward evaluation played an important role in determining the level of evaluation use. The case studies demonstrate that utilization increases when significant decisionmakers support evaluation and when the local community expects to be involved.

Williams' analysis of Suburb, West, sees the importance of the administrator's attitude in affecting evaluation use. He reported that "all acknowledged that the principal plays a key role" in determining utilization and that "a critical factor is the principal's openness and interest in the program." He interviewed two Suburb, West, principals and found them "using the evaluations and very enthusiastic about them." Similarly, Braskamp reported, "As one school board member put it, 'The superintendent made evaluation a top priority. It is working out quite well.'"

The importance of community expectations and attitudes toward evaluation is a recurring theme in these case studies. Suburb, West, provides a strong example. As Williams states, "Suburb, West, Unified

School District is a district with a citizenry that expects to be involved in decisionmaking." He illustrates the impact of community attitude with a description of one incident. At a particular school there was a lack of parent participation in an evaluation survey. "The parent committee members felt the low return rate was because parents didn't really believe that anyone would give their opinions any attention. Once the parent committee member reported that the data were indeed being used, the percentage of survey returns increased."

Local focus of evaluation. The case studies strongly suggest that utilization increases when evaluations are specifically designed to meet local needs. The evaluators we interviewed adapted TIERS data to provide answers to decisionmakers' questions. They also were ready to provide special services as the need arose. Willingness to tailor evaluations to meet local requirements strongly influenced the impact of evaluation data on decisions in these case studies.

Williams makes this point clearly: "Title I evaluation utilization can be greatly increased if a process that specifically addresses the decision needs of local school site advisory committees is developed and implemented." He points out that "Suburb, West, has had this need in mind as it designed, developed, and implemented an evaluation process almost entirely focused at the local school Title I advisory committee." Furthermore, Suburb, West, changed its testing schedule to allow advisory councils to use the evaluation data and report results in its April-May decisionmaking sessions. Individual schools collect data when it is best for them and get results back before they begin their planning.

Daillak notes that "over the years, Mr. Green has provided special services, on an as needed or as requested basis." He has provided analyses of issues raised by administrators and prepared evaluation plans for program proposals, for example. These special evaluation services have resulted in a great deal of utilization. Daillak cites an analysis of test score declines conducted by Green, at the request of district personnel, which resulted in grade level reorganization. Daillak attributes the success of utilization in Big City in part to Green's

timely response and sensitivity to local concerns. As he says, "Mr. Green did his best to respond promptly and usefully to administrative requests for information," and "he was responsive to teacher and school sensitivities."

Effective presentation of results. In each of the studies, the method used to present evaluation results, or the format in which they were presented, played an important part in gaining utilization. Graphics, monitoring systems, and narrative descriptions keynoted the methods that case study evaluators used to present results.

Williams describes the ways in which the evaluation data of Suburb, West, were reformatted to increase their usefulness to decisionmakers. Approximately 50 pages of computer printout were reduced to three pages presented in three graphic formats:

(1) "Tracking" particular groups of students as they progress through the grades;

(2) Showing the performance at each grade level for the various years involved; and

(3) In a profile which presents the gains realized from one year to the next for the various grade levels.

Small Town's evaluation team also converts its evaluation data into graphic formats for presentation. King says that Mrs. Foley "draws charts and graphs to show the results visually and to get information in usable form." According to King, Mrs. Foley finds these graphic formats "of value in making presentations." The Small Town Skills Charts present the results for each teacher's students using an objectives management system. King says of the charts, "The test results have been transformed into a highly explicit diagnostic tool for the Title I classroom, and the teachers are able to devise what amounts to an individual educational plan for each of their students."

Big City, too, has developed a unique format to present its evaluation results. The evaluation reports incorporate test score results and narrative evaluation critiques, which describe program strengths and weaknesses and give recommendations for program change. Daillak's evaluation of the narrative critiques indicates that they "appear to have gained a sympathetic hearing." In fact, the narrative critiques

"appeared to be more influential than the test score summaries in the evaluation reports. The narratives made sense to administrators and teachers, and seemed to carry weight."

Assistance in developing procedures for use. The case studies demonstrate that utilization increases when decisionmakers have specific steps to follow. Both Suburb, West, and Small Town had evaluators who developed procedures to help their decisionmakers learn how to use evaluation data. Mr. Leonard, evaluator in Suburb, West, believed that "school administrators, teachers, and parents who serve on Title I advisory committees often do not have group process skills and decisionmaking skills. They must be given assistance in how to read, analyze, and make decisions upon evaluation data." To this end Leonard has outlined a planning and decisionmaking process which lists the specific steps in sequence, the data that are available to inform the decisions, and the specific decisions that must be made by the district for reporting decisions. Over time the process has become increasingly familiar and useful to school advisory committees.

Small Town's evaluators have a slightly different approach. According to King, Mrs. Lopenia regularly encourages others to use evaluation data by giving them information and a related task or problem such as "suggesting changes in the program, planning next year's workshop, developing a dissemination plan, creating a meaningful attendance policy." In essence she is training Small Town's personnel to use data in making school decisions by giving them practice with the process.

Conclusions

Our review of the literature on Title I evaluation indicates a steadily growing awareness that evaluation utilization occurs. Moreover, the recent literature has begun to highlight the gradual incremental ways in which this utilization takes place. The SEA and LEA data we gathered in this study confirm and further amplify these prior findings. Both TIERS data and other types of Title I evaluation data are used at all decision levels. The Title I evaluation system does, indeed, have utility.

While we found relatively few examples of immediate, dramatic impact on decisionmaking, we uncovered strong evidence that evaluation data act as a contributing information source in the daily life cycle of projects and in incremental ways to more major program decisions. At the SEA level, evaluation data are used to monitor LEA compliance, to recognize both problem areas and exemplary programs, and to influence administrative and curricular actions. LEAs typically use Title I evaluation data to change attitudes and opinions toward Title I projects, to recognize situations requiring attention, and to contribute to decisions on administrative and curricular actions.

We also found that the kind of evaluation data used by LEAs varies with organization level. School boards, district advisory committees, and external agencies rely on TIERS data more extensively than other evaluation data. At the district administrative level, TIERS data are mixed equally with other Title I evaluation data developed by the district. At the building level, principals, coordinators, and others rely slightly more on TIERS than on other data. Finally, at the classroom level TIERS data are rarely used; instead, data more closely related to the instructional programs are preferred.

Since we had selected SEAs and LEAs thought to have high levels of evaluation use, the identification of factors associated with that use constituted a primary focus of the study. Analysis of the LEA case studies showed that six contextual variables affect evaluation's use. These factors are described below:

- Evaluator credibility. The reputation and credibility of the evaluators in each of the case studies was an important determinant of utility. While each of the evaluators achieved credibility in differing ways, they all were perceived as competent and trustworthy.
- Evaluator commitment to utilization. Credibility, while important, is not enough to insure evaluation utilization. The evaluator must also have a commitment to seeing that evaluation results are used by decisionmakers.

- Interest of decisionmakers and the community in evaluation. Utilization increases when both the decisionmakers and the community have a strong interest in evaluation and expect to be involved with it.
- A local focus to evaluation. Evaluation data are used when they provide information tailored to the needs and interests of the local school community. Utilization occurs when evaluators draw relevant information from TIERS data and when they conduct special evaluations to meet local requests.
- Effective presentation of results. Graphic, narrative, and non-technical modes of presentation increase the utility of evaluation data to local decisionmakers.
- Assistance in developing procedures for the use of evaluation data. Evaluation use increases when decisionmakers are assisted in understanding how they might use the evaluation data. Successful evaluators in our study typically provided detailed, step-by-step procedures to potential users.

The results of this study lead us to consider several recommendations for improving evaluation utility when Title I is replaced by Chapter I in the new law. Chapter I evaluation utility may be enhanced by continued technical assistance in reformatting TIERS-like results to meet LEA information needs. Moreover, SEA and LEA evaluation units should be encouraged to design a variety of local decision-focused evaluation strategies. In particular, locally designed evaluation procedures might provide information on the impact and costs of various materials and processes used within projects.

Many local and state agency personnel require guidance in developing procedures to follow when making decisions. It is not so much that administrators do not want to use relevant evaluation information as they typically do not know how to incorporate it into their decision processes. Finally, evaluators must become aware of the vital role their personal style plays in evaluation utilization. Training procedures for evaluators must highlight the evaluator's role and increase interpersonal skills.

REFERENCE NOTES FOR CHAPTER FOUR

1. Goettel, R. J., Kaplan, B. A., Orland, M. E., Forgione, P. D. Jr., & Huff, S. M. A study of the administration of the Elementary and Secondary Education Act (ESEA), Title I in eight states: Volume 5 -- A synthesis report. Syracuse, N.Y.: Syracuse Research Corporation, October 1977.

2. Forgione, P. D. Jr., Kaplan, B. A., & Orland, M. E. Evaluation of compensatory education programs: Problems, promising strategies and recent trends. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

3. David, J. L. Local uses of Title I evaluation. Palo Alto, Cal.: SRI International Educational Policy Research Center, July 1978.

4. Hamilton, G. E. Title I evaluation from the perspective of the local school. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

5. Boruch, R. F., & Cordrary, D. S. An appraisal of educational program evaluations: Federal, state and local agencies. Evanston, Ill.: Northwestern University, June 1980.

6. Kennedy, M., Neumann, W., & Apling, R. The role of evaluation and testing information in Title I programs. Cambridge, Mass.: The Huron Institute, September 1980.

7. Crowley, D. LEA usage of Title I evaluation data for decision making and program planning. State of Indiana: Division of Compensatory Education, Department of Public Instruction, March 1981.

8. Cichon, D. J., & Herr, G. Use of Title I evaluation: Results in an urban school district. Hampton, N.H.: Region I Title I Technical Assistance Center, RMC Research Corporation, July 1981.

9. Stecher, B. S., Alkin, M. C., & Flesher, G. Information use in elementary school decision making. Los Angeles, Cal.: University of California at Los Angeles, Center for the Study of Evaluation, December 1981 (CSE Report No. 160).

10. Alkin, M. C. Improving Title I technical assistance: An evaluation utilization perspective. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

11. Stonehill, R. M., & Groves, C. L. "U.S. Department of Education policies toward ESEA, Title I evaluation utility: Changes in attitudes, changes in platitudes," Educational Evaluation and Policy Analysis (in press).

12. Alkin, M. C., Stecher, B. M., & Geiger, F. J. The utility of Title I evaluation. Northridge, Cal.: Educational Evaluation Associates (in press).

CHAPTER FIVE

Changes in Evaluation Practice at Local and State Levels and the Possible Influences of TIERS

This chapter identifies the changes in Title I evaluation practice in state and local education agencies and the possible influence of federal evaluation reporting requirements, especially TIERS and technical assistance, on those changes. It attends especially to reporting and monitoring requirements made mandatory in 1979 by ED and made optional in 1981.

Evidence about change covers the period 1965-1981. Conclusions offered here are based on evidence derived from: (1) six site visits (two SEAs and four LEAs); (2) content analysis of evaluation material obtained from 12 state agencies; (3) previous statistical surveys of ESEA Title I evaluation practice; and (4) available reports from states receiving ESEA Title I refinement contracts. The focus of the report is on changes in the way certain activities are conducted (e.g., needs assessment and selection of students for Title I, estimating the effects of the program, and changes in capability of those responsible for evaluation). Our inferences about how much influence TIERS and TACs have had are weak because events during the period other than federal requirements also influenced changes.

The Needs of LEAs, the Needs of Students, and Program Planning

Front-end evaluation practice has changed for the better in LEAs and SEAs. Not all changes are plausibly attributable to TIERS and related activity; some are.

Needs assessment.

The years 1965-70 were characterized by very little rigorous work by LEAs and SEAs in identifying needy school districts. Since then, the data being generated have become more systematic, explicit, and verifiable, at least in principle.

The authors of this chapter are Robert F. Boruch, David S. Cordray, Georgine M. Pion, Bruce D. Spencer, William M. K. Trochim, and John Wick.

At the state level, the evidence that early work was dismal is reflected in case studies. A Title I director in State A, for instance, maintained that "about all an LEA had to do was declare the need for a program in industrial arts, reading, and so on, and money would be provided." The state now requires explicit justification of needs based partly on test scores. The improvement in State D is more recent. Need still hinges on "what is liked best" in the district. But in 1981, a concerted effort was made in consultation with the regional TAC to establish a framework for ranking needs better, based on educational theory and research, needs assessment data, and previous test data.

Statistical evidence has been generated by Gamel, Tallmadge, Wood, and Binkley, (Note 1), by Wargo, Tallmadge, Michaels, Lipe, and Morris (Note 2), and by others which suggests that between 1971 and 1974 the proportion of states that provided evidence on needs assessment was about 30 percent and stable. During the same period, however, the nature of assessments changed: more standardized tests appear to have been used to document needs. All states now must be explicit in reporting how its LEA needs are assessed--bringing the nominal rate of states presenting evidence to 100 percent. But no strong evidence on quality of this reporting is available. The best available national evidence, from the Sustaining Effects Study in 1976-77 (Wang, 1978), suggests that there are still needs not being met, if the basis for judging needs is a high fraction of poor achievers not involved in LEA Title I programs. TACs have informed the needs assessment process, primarily but not exclusively at the SEA level.

We believe the change is probably partly attributable to the federal stress on reporting, though not specifically to TIERS. As expected of a large LEA in State D, respondents reported that neither TIERS nor TACs influenced small changes that did occur. Other probable influences include studies of the criteria for determining needs, such as the Compensatory Education Study (1977) of the National Institute of Education (e.g., income level versus test scores), OE's studies during the 1970s, and audits.

Student selection and local practice.

At the local education level, the change in documentation of children's needs is reflected in monitoring and in the contents of

reports containing specific information about how children are selected for Title I programs. The change is not uniform.

To judge from interviews in State A, this evidence includes specification of: grade levels and fraction of children likely to need assistance based on test scores; the tests used in making judgments; cutoff points for admission to programs; and other data, all submitted as part of applications for Title I funds. The state Title I monitor examines these to verify that indeed the children exist and achieved the test scores advertised, that certain cutoff points are used to determine eligibility, that exceptions and dropouts are documented. In principle, the information is verifiable, but special studies would have to obtain the corroborative evidence. In practice, the quality of information varies within state. For example, the change in State A involved moving from a situation in which "a teacher's recommendation only" was sufficient for installing a child in a Title I classroom to explicit ranking of children on the basis of a verifiable achievement test score and judgments of teachers and independent tests where the single test score is suspect. In one of the largest LEAs in State D on the other hand, tests and teacher ratings have been used traditionally. In 1980-81, the SEA imposed a point system requiring that final student rankings be considerably more explicit, through the use of teacher ratings, multiple prior test scores, and other information.

Shifts that have occurred do not appear to have been influenced directly by federal reporting requirements but do appear to have been influenced by policy stress that the neediest be served and by audits that focused on needs. Respondents at large LEAs in States A, B, and D said that in assessing children's needs they are left alone by the SEA and do not get much assistance from the TAC. We do not know how much small districts depend on TIERS, TACs, or other such outside systems to guide the process, however.

Number of students served.

The estimates of the number of children that result from more systematic reporting are, of course, critical at the national level, and those numbers appear to have improved. This does not mean counts are unambiguous.

During the early years of Title I, counts of students were not often interpretable. Surveys of Gamel et al. (Note 1) over 1969-74 lead to a conclusion that duplicated counts, state mixing of duplicated and unduplicated counts, ambiguity in counting rules, and other problems in state Title I reports made a total count of students very suspicious. Between 1971 and 1974, only six to nine state reports were of sufficiently high quality and reported in a sufficiently consistent format to consider using them in national estimates. The 1980 data reported to ED are unduplicated counts and less ambiguous, to judge from the Department's FY 1981 Annual Evaluation Report on Educational Programs.

Program planning and objectives.

There is a little evidence that progam planning and specification of objectives have improved. The improvement that has occurred may be attributable to stress on evaluation, which in concrete form includes TIERS.

In site visits to a large LEA in State A for instance, we were told that early program planning and objective setting were erratic processes at best. LEAs were not required by the state to specify objectives in proposals to the state. In 1973, the Title I LEA supervisor developed a more systematic, management oriented approach, and in 1980 this became policy for all districts. Systems with similar purpose, clarifying what is expected of children and teachers, have also been created in the large LEAs visited in States C and D.

In State A's district, planning time is set aside each week for Title I and classroom teachers to discuss objectives and related progress. The entries in the reporting system for recording objectives (Individual Education Plans) are examined by the state monitor who regularly visits Title I classrooms. Evaluation personnel also participate to the extent of refining objectives and making statements of objectives clear.

Published descriptions of similar work are available. Rost (1980), for instance, reported on the use of a four-part planning, evaluation, and dissemination system that attempts to facilitate the use of evaluative information for local planning of Title I. An Item Documentation manual was developed to provide teachers with information on what each test item was designed to measure. By reporting test

results on an item-by-item basis, the teacher can use the manual to determine readily which skills should be focused on for each child. A School Report synthesizes evaluation data in a manner that makes it conducive to planning. The Resource Allocation Plan, based on Title I needs assessment data, includes a summary of the school's relative need for supplemental assistance. Finally, the Title I Impact Review is conducted. Here, the instructional team reviews the evaluation data with the evaluation staff--giving reasons for specific outcomes based on events occurring during the school year. The direct effect of TIERS on all this cannot be established from the evidence at hand.

Indirect evidence regarding interest in conducting evaluations relevant to program improvement comes from a survey conducted by the National Center for Education Statistics (NCES) and from TAC yearly reports. The NCES survey (1980) shows that an estimated 44 percent of the districts desired help in devising evaluations for program improvement. The TAC reports for Regions VIII, IX, and X document that for recent years (1980-1981) the type of requests they receive for assistance changed from being predominantly focused on TIERS in the early years (1976-1979) to general evaluation issues.

Achievement Testing and Implementing the Evaluation Models and Reporting System

The raw material for judging the effect of programs on children comprises achievement tests, and for that reason LEA and SEA attention to their quality is important. The models serve as a framework for making inferences from such tests, and thus their implementation is important, too.

Changes in quality control.

The quality of evaluation evidence has improved over the past 10 years, but the changes are not uniform.

In the 1981 ED report cited earlier, for the first time, evaluation evidence from TIERS is presented for all 50 states and the District of Columbia. ED is cautious about these data, pointing out that they are based on practices which may contain errors. The discussion presented here assesses the sources and types of errors that are likely to influence these data.

Many SEAs, through Title I state refinement contracts and/or TACs, have introduced procedures for data editing and quality control. These efforts vary from state to state in a number of ways. Some states have only recently used monitoring practice to understand errors in the data reported by LEAs; others began this quality control earlier and have initiated changes that have reduced error rates. For example, the SEA in State H found that for FY 80 TIERS data, roughly 40 percent of the pretests and 20 percent of posttests for reading and mathematics achievement were administered outside the appropriate norm dates. In general, less than 30 percent of the LEA reports that were reviewed had no errors or easily corrected errors. The remainder were deficient in one or more areas.

State C has opted for a probability sample of LEAs reporting each year. For FY 1980 reporting, the SEA's quality control procedures identified 15 percent (we estimate) of the LEA reports as containing errors that were sufficient to warrant their data being excluded from the state aggregation.

Other states are further along in quality control (e.g., States A, E, F, and G). Sources of errors have been identified in early pilot tests (e.g., State E), materials and workshops have been developed for training purposes, and some show evidence of reduction in errors (e.g., State F). For example, State F's SEA reported in 1976-77 that 40 percent or more of the LEAs made errors in testing, student selection, inappropriate tests, etc. During the following year (FY 1978), most of these errors were reduced by at least one half or completely avoided. Efforts in States B and P show similar increases in quality that can be attributed to deliberate efforts to edit and check appropriateness of evaluation procedures. The data are not perfect, but experience with TIERS coupled with quality control procedures have reduced some of the major errors. A few of our respondents suggested that before these quality control procedures were installed, the practices underlying the data reported by LEAs were not often questioned, much less examined systematically.

There is, of course, substantial variability in quality from one LEA to the next. State H reports that the smallest incidence of inappropriate test administration occurred in districts with large Title I awards (roughly 90 percent of the LEAs in this category used appropriate administration periods), and the highest occurrence of errors in test administration were for very small LEAs (i.e., small Title I awards). The data suggest a linear relation between size of award and quality of practice in this area.

The data from annual testing cycles are considered to be less subject to biases than data derived from fall-spring administration, and indeed in 1980 about 30 percent of those tested are on an annual testing cycle. The annual cycle raises the possibility of a different kind of bias, however. For instance, State H expects at least 66 percent of enrollees to be tested, which is possible in fall-spring testing. With annual testing, however, student participation in testing falls below this level because of student mobility and other problems.

In addition to test data, non-test information has been systematically examined in a few states. Again, State H reports that 40 percent of the LEA reports contained errors in specifying project characteristics. Little else is known about the quality of these data. Site visits to large LEAs in States A, B, and D suggest that some LEAs are making concerted efforts to monitor the quality of information as it is gathered to assure that it is at least consistently recorded. In one site, the evaluator verifies the information by sending it to the classroom teacher to check its accuracy. In another site, the resource teachers monitor and spot-check the data monthly.

The use of the models.

The quality of evidence on the effects of Title I has improved but not uniformly.

The adoption of models by states is taken here as an indirect indicator of improvement--assuming that the models are technically sound in the first place. Other evidence is more direct.

An increase in uniformity in evidence, apart from quality, can be educed from a report by Anderson, Johnson, Fishbein, Stonehill, and

Burns (1978). At most, 10 of the states produced data sufficient for aggregation in 1977. By 1978, the number had increased to about 20, and by 1980 all states were providing data for national aggregation. As one would expect, opinion about quality of data reflect this shift. For instance, nearly 85 percent of Title I coordinators responding to Kearns' survey (Note 3) maintained that TIERS results are likely to be more accurate than data collected previously. This figure is substantially higher than would be inferred from earlier surveys by the GAO (U. S. Comptroller General, 1977) dealing with evaluation quality before 1975.

One indicator of local and state improvement in estimating the effect of Title I programs is the extent to which LEAs adopt the evalution models developed for Title I by the federal government. Assuming that these models are indeed appropriate, they may be implemented well or poorly. Information on implementation comes from several sources.

The Region I TAC, for example, traced implementation of the models over three years. The following table gives the fraction of LEAs in the region "attempting" to implement the models in each year and the proportion doing a "rigorous" job.

Table 5.1

Status of Model Implementation by LEAs in Region I

	1976-77	1977-78	1978-79
Attempted	52%	76%	100%
Rigorous	22	37	57

Six states and 838 LEAs are covered by the region. It is clear that rigorous implementation has increased. It seems reasonable to attribute the increase to ED/TAC efforts to encourage implementation.

The following table, taken from Anderson et al. and NCES gives the proportion of LEAs implementing the models to a greater or lesser degree in each of the Title I regions (in parentheses):

Table 5.2

Percent of LEAs Implementing Models

Region	1976-77	1977-78	1978-79
Atlantic (I, II)	36%	43%	73%
Great Lakes & Plains (V, VII, VIII	18	45	57
Southeast (III, IV)	15	34	52
West and Southwest (VI, IX, X)	27	33	49
Overall	23%	40%	57%

When the statistics for 1978-79 are broken down by school district, we estimate that about 77 percent of the large districts (10,000 or more students) implemented the models, about 65 percent of districts with between 2500 and 10,000 did so, and about half the small districts did so. Closer inspection of implementation data suggests that there is substantial regional variation in implementation rate in any given year (e.g., from less than 1 percent to over 80 percent for 1977) and that there are notable regional differences in rate of change from one year to the next. Supplementary statistics also underscore an earlier finding of Boruch, Cordray, Pion, and Leviton (1981): states vary considerably in their ability and willingness to evaluate. In particular, there are two distinct classes of states -- those implementing TIERS fully and those that are considerably more cautious or less able to adopt the conventional reporting model. At one extreme, a state implemented the system reasonably quickly, and moreover the state's legislature then mandated use of the models as part of the state compensatory education program. On the other hand, some states were cautious, in that they pilot tested the reporting system prior to full scale implementation, (e.g., State I). Others chose a voluntary approach, permitting LEAs to adopt the system slowly over a longer time period.11/

11/ The reasons for caution vary. Some states took pains to assure that the Title I models were compatible with their systems for evaluation of state compensatory education programs. Others, such as State K, had recently devised their own reporting system and were reluctant to modify it to conform to TIERS requirements.

The improvement at the state level seems attributable in part to ED stress on evaluation, on resources expended in special studies, and on technical assistance. Reports submitted by states are influential only insofar as feedback on flaws is provided in the interest of improvement. But the federal influence is virtually inseparable from state initiatives, in that the Title I reporting system did develop with considerable state involvement (see Kearns, Note 3, for example). In at least one case, the TAC influence was clear: a state that merely distributed technical material on models later discovered substantial imperfections in data and then leaned heavily on the TAC for guidance. The federal influence is clear, in that it is unlikely that all states would have implemented the models without the federal initiative.

General effects on reporting.

Between 1965 and 1980, LEA reports, and the SEA and federal reports based on them, changed dramatically from essentially public relations documents to reports that rely at least partly on more reliable and verifiable data.

The early evidence for poor quality of reports stems from case studies by McLaughlin (1975). She observes, for instance, that the 1966-67 reports contained a great deal of rhetoric and gratuitous cheerleading, Title I programs being "an educational revolution," "vibrant," "reflective," etc., with claims that the evaluations showed the effect of programs to be "unquestionable and dramatic." But reports presented little evidence meeting reasonable standards of evidence. Wider scope reviews of state Title I reports by Gamel et al., (Note 1) and Wargo et al. (Note 2) suggest strongly that during 1969-74 less than 75 percent of the states provided adequate information on the way they selected samples of students discussed in reports; the reports of no more than nine states were sufficient to even consider combining to obtain national estimates of Title I performance; less than 70 percent contained any real information on the needs of children; and most did not provide information with sufficient clarity to even count children receiving Title I services.

Both the contents of required reports and the quality of information have changed. For instance, narrative reports that included statistics were provided to the federal government up to about 1979. In 1980, the first formal TIERS report was submitted covering only numerical data. The data consist of numbers of students, gains in standard measurement units, grade levels, and other information that is also being used at state levels to monitor student progress. This is a reduction of paperwork burden on states, since no narrative information is required. But it also means that the reports, when considered by themselves, become hard to interpret.

The statistical nature of change in quality can be understood partly from the reviews by Wargo et al. (Note 2) of state Title I reports from 1969 to 1974. The period 1969-1970 was characterized by no detectable change. But ratings of quality for 1971-74 show notable positive changes in readability of reports, in reporting about cognitive achievement of students, in data adequacy, and in policy relevance of the reports (Gamel et al., Note 1). For instance, the fraction of reports containing information on cognitive achievement that was probably biased (by statistical artifacts, sampling bias, etc.) declined from 45 percent in 1971 to 23 percent in 1974. The fraction of reports based on samples that were clearly biased declined from 40 percent to 17 percent. This does not mean that reporting is perfect, of course. Despite the decline in "clearly biased" samples, the fraction that are rated as being "probably biased" doubled.

Capabilities, Knowledgeability, and Independence of Evaluation Staff

Reliance on more expert staff.

There has been a change in views about what is reasonable to expect of staff in regard to evaluation and a change in reliance on more expert staff.

The expectation in some states now is that it is not possible for many staff members in LEAs to become experts in evaluation and that even where experts can be hired, high staff turnover requires that alternative

systems have to be found to assure adequate data. 12/ In State A, for instance, as elsewhere, principals, teachers, and others may be responsible for providing data in an evaluation but often do not have the resources or incentives to do a good or even adequate job. Rather, the state's objective is to get concerns across (e.g., that tests have to be selected carefully so as to be material to courses and must be administered carefully and at the right level). The recognition that it is not sensible to expect administrative fiat to result in perfect data is reflected in the creation of a quality control system a few years ago for the data submitted to State A, and similar systems have been set up for similar reasons in many states.

Where exceptions occur, they are most likely to be in the large LEAs, for instance, in States A, B, C, and D. In State A, for example, two evaluation staff members had training beyond the Master's degree, were more technically sophisticated than evaluation staff at the SEA level, and were active in presenting papers in professional forums, in obtaining grants, and so on. They rely for technical assistance on faculty at a nearby university. The large LEAs visited in this study and earlier ones confirm that they may have staffers with substantial postgraduate training.13/

The changes here were alleged in State A to have run parallel to changes in the caliber of Title I teachers over the past 10 years. "At the beginning. . .anybody could do it -- pre-retirees, and so on. Now the state requires x hours of training for certification, and the Title I teacher is an expert."

12/ LEA staff turnover prompted one TAC to speculate that the models would probably never be fully implemented without continual training of new personnel. A state that initiated an evaluation improvement program expressed much the same view.

13/ We were told by one TAC representative that states which they served were interestd in using technically sophisticated LEA staff as a formal resource for remaining LEAs, in the event that TACs are terminated.

That there are limits on the extent to which LEA and SEA staff can routinely produce evaluative data has also become very clear. Turnover alone is a formidable limit (e.g., about a third of LEA Title I staff are replaced each year in one state), and partly as a consequence, maintaining quality of testing and use of models will be a persistent problem. The federal resources that have been used to expand those limits include TACs, of course. The development of formal data quality control programs, for example in State A, is important since technical burdens are reduced through such systems.

Within the last year, the concerns of program coordinators appear to have shifted in a way that threatens the progress achieved. In particular, it is possible that budget cuts will lead to cuts in evaluation and that good people will leave or be relocated or terminated. The possible, and we believe probable, consequences are, first, lower quality control, followed by gradual decline among states that are emergent, and second, no progress and probably decline among states that have made no progress so far. We do not believe the progress that has been made, such as it is, can be maintained without state and federal encouragement.

Knowledgeability.

What is understood about estimating the effect of programs at the local and state levels is considerably less naive now than it has been in the past. Further, the quality of attempts to get at effects has improved. The wild claims about effectiveness of programs at the local, state, and federal levels during the late 1960s, evident in early congressional hearings, press releases, and so on, has decreased. Understanding the circumstances under which impact estimates are possible has become more sophisticated. For instance, State A stresses now that such evaluation is often unreasonable at the local level because the number of program participants is usually small and measures of achievement are imperfect.

There are more cautious assertions in State A that impact at state level can be estimated reliably. The caution is reflected as well in recognizing that influences other than Title I exist simultaneously and that it is impossible to separate these from Title I. For some

SEA administrators, it is also reflected in the idea that Title I models for evaluation are not perfect but "they are all we have." The recognition has been discussed in public professional forums by Kearns of the Arizona Department of Education (Note 3) among others, as well as privately, and it extends to thoughtful skepticism about estimates of program effect at the national level.

Independence and candor.

At the local and state levels there are a few illustrations of efforts to foster independence and honesty of evaluations. One school district, for example, is creating a remarkable policy of independence of its evaluation units so as to assure that evaluators can report honestly and avoid unnecessary pressures to "adjust" data. In State B, findings by a local evaluation unit that children were being "fed" questions to improve achievement test scores resulted in attempted coverups and controversy, and this led to reorganization of reporting responsibility. Testing and evaluation units were made administratively independent of the program to reduce self-serving interests on either side. The LEA in State D encountered related problems during the 1970s, and evaluation was reorganized to avoid unseemly influence by the LEA's director of federal programs. Such changes are a matter of state and local initiative rather than federal action.

Going Beyond the Reporting Requirements

Where change is discernible.

Some, not all, SEAs and LEAs use the federal stress on evaluation, including reporting requirements, to develop reporting systems more useful to themselves and as leverage to mount studies that improve programs. The increase in this activity since the late 1960s and early 1970s is remarkable.

Some LEAs and SEAs do indeed build on federal requirements to understand more fully how to improve evaluations and programs. A survey of LEAs, conducted for the recent mandated of educational evaluations (Boruch et al, 1981) showed that except for small districts at least three quarters of LEAs went beyond federal requirements during 1979-80 (i.e., supplemented federal reports with information of other kinds).

In State A, for instance, the chief state school officer initiated statewide statistical studies to understand the influence of pullout, class size, and other factors on special achievement in Title I programs. In site D, a special study of the counseling component of Title I services has been mounted, augmenting existing information with systematic behavioral assessments of children through tests, parent and teacher ratings, and other data. The staff also conduct studies of test reliability, relevance, and validity, and have mounted a longitudinal study of Title I children. Where local capability is high, such efforts do not often involve TACs; where it is not, TACs are exploited at times.

Submission of projects to the JDRP is relevant too. State A, for instance, has had four projects approved. The Title I director and monitor maintained during interviews that the data accumulated as a result of federal reporting requirements and the requirements themselves were influential in provoking interest in obtaining approval and guiding collection of evidence. In particular, leaving the TIERS Model A in place facilitated the task of getting data necessary for JDRP approval. The activity within State A's largest school district corroborates this. The staff are involved in special studies of achievement as a function of time on task and in instruction, between achievement and Title I attendance, and other studies in addition to work submitted to the JDRP. TACs are used as a technical resource (e.g., to critique submissions to JDRP). The "going beyond" extends to the monitoring process as well, according to Title I officials in the state.

State refinement grants have played a limited role in this, simply because the number and size of such grants have been limited. Grants that have influenced quality control are discussed in the earlier section on the topic. Examples aside from these include one state's feasibility study of microcomputers to facilitate Title I data collection and decision making.

Where the stress on evaluation has had little discernible influence. The effect of TIERS and more generally the effects of stressing evaluation appear to have been negligible in fostering high quality tests of costs and benefits of alternative program variations.

Nothing in the legislative mandate requires that such tests should be undertaken. Still, the history of Title I and the stress on innovation suggest that an ultimate goal is to get at what is best at a specific cost; and that is absent. Comparisions that are made are usually weak. Given the variety of approaches and the inventiveness of some LEA and SEA people, the absence of systematic tests is a disappointment.

To cite a specific example, consider that in State A funds are available by the SEA to support innovations at the LEA level. About $200,000 per year are disbursed in a kind of competitive grant program to Title I teachers. The $5000 average grants are made to teachers for work during the summer to improve management systems and mathematical skills, to couple skill requirements and objectives with materials, to develop materials that can assist parents in teaching at home, and so on. Typically, grantees reapply a second year to complete projects and often, according to state officials, LEAs adopt the inventions of their colleagues in other LEAs. None of this, however, is evaluated systematically. The lack of evaluation may be due to the view that small projects warrant no more than informal evaluation or may be due to indifference or inability.

More general evidence for the lack of change in this area comes from statistical surveys of Gamel et al. (Note 1), Wargo et al., (Note 2), and others. They found that the use of certain evaluation designs, such as before-after measurement, remained very stable during 1969-74, and designs involving formal experimental and control groups actually declined from a very low level to virtually zero.

Exceptions to this trend depend heavily on availability of skills and resources. So, for instance, staff of one large urban school system and a nearby state university have collaborated on development of school based strategies for treating children with learning and behavioral difficulties in Title I preschool programs. Moreover, they have assessed the intervention using a randomized field test, blind ratings, and placebo

controls to determine that the intervention performs as advertised in identifying problem children and in improving their adjustment. Well designed field tests of this kind are the exception rather than the rule.

REFERENCE NOTES FOR CHAPTER FIVE

1. Gamel, N. N., Tallmadge, G. K., Wood, C. T., & Binkley, J. L. State ESEA Title I reports: Review and analysis of past reports and development of a model reporting system and format. Mountain View, Cal.: RMC Research Corporation, October 1975 (RMC Report UR-294).

2. Wargo, M. J., Tallmadge, G. K., Michaels, D. D., Lipe, D., & Morris, J. J. ESEA Title I: A reanalysis and synthesis of evaluation data from fiscal year 1965 through 1970. Palo Alto, Cal.: American Institutes for Research, March 1972 (AIR-27400-3/72-FR).

3. Kearns, D. R. Impact of the USOE Title I evaluation and reporting system: An SEA perspective. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, Cal., April 1979.

CHAPTER SIX

The View of TIERS from Washington

This chapter of the report examines the effectiveness of TIERS in meeting the needs of ED and Congress. The effectiveness of TIERS at the federal level has been considered from three perspectives:

- As a reflection of the federal implementation of a specific legislative mandate;
- As an evaluation system producing information needed by ED and Congress; and
- As a model for grantee evaluation and reporting systems used in other programs.

These perspectives were chosen because they permit a comprehensive view to be taken of the federal interests in and uses of TIERS and its associated program of technical assistance.

To place this inquiry in its proper context, it is important to state a perception that may be obvious to many educators and evaluators who have worked with TIERS at state and local levels. That is simply that TIERS embodies an astonishingly ambitious idea. The first part of this undertaking was the mandate to federal officials to design a set of plans, or models, capable of guiding the local collection of certain data that are particularly complex, politically sensitive, and hard to measure (i.e., achievement data). The second important aspect of the system is the intent to use the information produced through implementation of these models at local, state, and national levels in addressing two very different purposes, which are, first, to assess the relative success of the Title I program in meeting its statutory goals and, second, to serve as a roadmap for the design of improvements in service delivery.

The ambitious nature of TIERS is highlighted by two additional phenomena. First, Congress conceived that this system would be implemented in 14,000 separate administrative units, or school systems, each of which is divided into still more sub-units, or Title I schools. Second, the object being measured (i.e., student achievement) is the

---

The author of this chapter is Elizabeth R. Reisner.

product of the efforts of one of the largest and most politically powerful occupational groups in the nation (i.e., teachers).

Yet, while the idea of TIERS was grandiose in scope, so were the resources provided to enable OE and then ED to implement the mandate. From the point of view of time, money, and personnel, the agency was richly blessed. 14/ Did the agency's performance meet the challenge of its assignment? Were these sizeable resources well invested by the Executive Branch? The remainder of this chapter will explore these and other questions.

Methods of Inquiry

Data on which this study is based were collected using three techniques. First, relevant articles, books, and other publications were obtained and reviewed. Second, correspondence between OE and the congressional committees also was obtained and reviewed. Finally, interviews were conducted with ED staff and staff of the congressional committees. Each of these data collection procedures is discussed more fully below.

Literature review. The purpose of the literature review was to determine how other researchers and political observers had looked at the issues of concern to this study and what they had found. No analyses, either published or unpublished, could be identified that addressed the precise subject of the study. A number of other writings were identified, however, that discussed certain aspects of the issues being reviewed. These studies are cited in the sections of this chapter to which they pertain.

Review of correspondence. Our reasons for reviewing the correspondence between OE and Congress were to answer two broad questions. First, in what ways were the congressional committees involved in the implementation of the requirements of Section 183, and what was the effect of that involvement? Second, what changes in TIERS did the committees request through correspondence (as contrasted with legislative amendment), and to what extent were those changes implemented?

14/ In addition to the seven years of time used to implement the 1974 requirements, the agency was permitted in some years to spend as much as $8.2 million on TIERS and the associated program of technical assistance. Until recent personnel cuts, ED has employed five doctoral-level professionals to spend most of their collective work time on the implementation of TIERS and administration of the TACs.

To answer these questions, we assembled a correspondence file that included those letters and reports exchanged between the Executive Branch and Congress on the topic of TIERS. We identified nine letters sent by OE to Congress during the period from May 22, 1974, to August 14, 1978. 15/ We were able to obtain copies of all of these letters. In addition, five status reports on TIERS implementation were transmitted to Congress during this time period, either in the form of formal reports or as written testimony transmitted during hearings. 16/ This number does not include the annual OE evaluation reports submitted to Congress during this period, each of which contained chapters on Title I. We identified ten letters sent by Members of Congress or congressional staff to OE or ED. Of these we obtained eight.

Interviews. Personal interviews with 19 persons were conducted in order to learn more about information needs for Title I (and later Chapter I) experienced by managers and staff of ED and by Members of Congress and their staffs. Interviews were conducted by the author using a written interview guide. Although some questions were asked of more than one respondent, the overall set of questions was different for each respondent. Each interview required 30 minutes to two hours of the respondent's time.

Federal Implementation of the Legislative Mandate for TIERS

With its submission to Congress of the FY 1981 Annual Evaluation Report on Education Programs, ED provided Congress and the public with the first analysis of Title I achievement data collected from all 50 states using TIERS. More than seven years of effort had been required to implement this system. According to one senior congressional aide, that period represented a remarkably slow pace, lengthened mainly by governmental delay (Jennings, Note 1). According to a former OE evaluation manager, however, those seven years had been the minimum period

---

15/ In a number of instances, duplicate letters were sent by OE to several Members of Congress. Only the original version of each of the letters was counted in determining these numbers.

16/ One of the nine OE letters consisted mainly of a formal status report on TIERS and is therefore reflected in the count of status reports.

necessary to complete the technical and political tasks needed to achieve full, nationwide implementation of the system (Evans, Note 2). To assess the performance of OE and, later, ED in this regard, it is necessary to identify, first, the tasks that were set before the agency by the 1974 amendments and, second, the steps that the agency took to implement those tasks. Using this information, it should be possible to find answers to questions such as the following:

- Were the requirements of Section 151 and, later, Section 183 fully implemented by OE and ED?
- What aspects of the implementation presented the greatest problems for the agency?

Steps required to implement the mandate. Examination of the provisions of the 1974 mandate indicates that implementation of the key provisions had to occur sequentially rather than concurrently. This pattern can be seen through an examination of the basic framework of requirements contained in Section 183 and in the other Title I provisions related to evaluation.

The first step in this sequence is the federal decision on what local project components will be evaluated and what success criteria will be set for each component. This step fulfills the requirement for the development of evaluation standards, as stated in Section 183(b)(1). The next step is to develop a set of models, or plans, for measuring local project success in implementing each of these components, as required by Section 183(d). Part of the model development process is specifying techniques and methodologies acceptable for local evaluations using each model, as indicated in Section 183(f). Once the models are developed, technical assistance, as required under Section 183(e), must be made available to those responsible for implementing the models in actual programmatic settings. The first task of this technical assistance is to impart an understanding of the purposes and potential uses of the models. Then, when the models are understood, SEAs can begin to develop sampling plans and evaluation schedules on the basis of OE guidance, as required by Section 183(b)(2). When the samples and schedules are ready, LEAs can be required to begin reporting on the evaluative data collected in

accordance with the requirements of Section 124(g). These data will be aggregated by SEAs and presented in state evaluation reports to the Secretary, as mandated under Section 172(1). At the federal level, state data will be aggregated to yield a national picture of the characteristics and effects of Title I, as described in Section 183(g).

In addition to using TIERS data for reporting program results from the local to state to federal level, the data are also to be used at the same three levels for educational improvement purposes. For example, Section 124(g)(3) requires that "the results of the evaluations will be utilized in planning for and improving projects and activities carried out under this title in subsequent years." General directions are given to SEAs, in the context of technical assistance requirements in Section 166, pertaining to the use of data from local evaluations and other sources for program improvement purposes. More specific requirements of this type are to be implemented at the federal level, as indicated in Section 183(h), in the use of evaluation data for program improvement, especially data identifying exemplary programs.

The development of TIERS. Although the framework for federal evaluation responsibility appears to be relatively straightforward as described in the law, the actual implementation of the legal framework has presented a number of problems not apparent in the statute itself. One early problem was skepticism as to whether the system could work, particularly whether a structure could be devised for local collection of data that would yield information valuable for decisionmaking at all three levels of Title I governance. To understand the character and the effects of this skepticism, it is necessary to review some of the events contributing to the development of TIERS. (See Appendix B for a chronology of communications between Congress and the Executive Branch regarding the development of TIERS.)

According to one senior congressional aide (Jennings, Note 1), the Veyzey amendment had been proposed and endorsed as a result of congressional dissatisfaction with the quality of evaluative information produced through OE's national studies of Title I. This aide stated that many members of Congress who had taken an interest in Title I feared that

Title I evaluation results had been shaped by President Richard Nixon's White House to yield intentionally negative findings about the effectiveness of this Great Society program. This perception was reinforced by President Nixon's 1970 message to Congress on educational reform, in which he stated that "the best available evidence indicates that most of the compensatory education programs have not measurably helped poor children catch up." With this rationale he had argued against increases in funding for Title I (cited in McLaughlin, 1975, p. 65). The Veyzey amendment was seen by many Members of Congress as an avenue for generating national Title I information that was untainted by prevailing political preferences (Jennings, Note 1). In a similar vein Christopher T. Cross, former legislative aide to the ranking minority member of the House Education and Labor Committee, has noted that congressional interest in the Veyzey amendment reflected dissatisfaction with the poor quality of Title I information available at the time of the 1974 reauthorization and a determination to improve the quality and quantity of Title I data available for the next reauthorization (1979).

OE officials saw serious problems with the Veyzey plan even prior to its enactment, however. According to testimony of OE Commissioner John Ottina before the House Education and Labor Committee prior to congressional approval of the Veyzey amendment, OE had doubts that national evaluation standards could be set in Washington that would be suitable for local Title I evaluation needs nationwide (Note 3). Despite OE's doubts, however, the agency responded to the enactment of the Veyzey amendment by revising its Title I evaluation plans and initiating the development of "model Title I reporting systems. . . which will suggest the appropriate achievement test-analysis matches, some analytical techniques to facilitate interpretation of the data, and reporting formats which will yield comparable data on a state-to-state basis" (U.S. Office of Education, Note 4).

Even with these early steps, however, concern was expressed by at least one senior legislative aide regarding OE's approach to the development of Title I evaluation standards. In February 1975, Christopher Cross wrote to the OE Commissioner, "It is not at all clear from

the description [that you sent earlier] how much emphasis you will place on the mandatory use of those standards and what your schedule is for doing so" (Note 5). OE's response to this concern was stated in a March 1975 letter from John W. Evans, OE evaluation director. While indicating concurrence with the Committee's interest in improving the quality and comparability of local Title I evaluation practices, Evans said, "Given the diversity of school procedures and the great resistance to Federal prescriptions in the data collection area, I personally believe it may never be possible to lay down sufficiently specific evaluation standards for local projects such that all the data they collect can be meaningfully aggregated" (Note 6).

This simmering dispute was brought to a head in an October 1975 letter from the House Education and Labor Committee signed by both Chairman Carl D. Perkins (D.-Ky.) and the ranking Republican on the Committee, Albert H. Quie of Minnesota (Note 7). In their letter they criticized the OE decision to spend 75 percent of the Title I evaluation funds on the national longitudinal study of Title I sustaining effects and urged that funds be shifted to the provision of technical assistance to SEAs in implementing Title I evaluation models. They said, "It appears as though the emphasis is very strongly on a national evaluation and that the questions of how technical assistance is to be provided to State and local educational agencies and whether or not the models for evaluation will be required have not been given adequate attention and thought."

OE's response to this critique was prompt. The study of Title I sustaining effects was dramatically reduced in scope and expense (from 74 to 24 percent of expected funds for Title I evaluation over the period from FY 1975 through FY 1979), and technical assistance to SEAs was increased from 5 percent to 65 percent of the expected five year evaluation budget (T.H. Bell, Note 8). Shortly after these plans were laid out to Congress in February 1976, OE announced to SEAs the availability of several manuals for implementation of standard evaluation procedures and asked SEAs to plan to attend regional workshops on Title I evaluation in the fall (Note 9). The first contracts to operate TACs were awarded at about the time of the first regional workshops.

Though these steps met the interests expressed by Congressmen Perkins and Quie, they provoked opposition from a number of Chief State School Officers, as indicated in a September 1976 letter signed by the school chiefs of Alaska, Hawaii, Idaho, Montana, Oregon, and Washington (Note 10) and a letter from the New York Commissioner of Education (Note 11). These letters asked OE to slow down its development of mandatory Title I evaluation requirements. OE's response to this concern was to try to address and resolve potential implementation problems while state and local adoption of TIERS was still voluntary. To achieve these objectives, OE expanded its outreach and coordination efforts through increased state visits to discuss TIERS. It also sponsored TAC workshops and initiated publication of a regular TIERS newsletter.

By mid-1977 the education committees in Congress were beginning to consider revisions to Title I that would be enacted in the upcoming reauthorization of ESEA. In questions submitted to OE in July 1977 by Congressman Perkins, the first hints were given that members of the House Education and Labor Committee had doubts about the assumptions underlying TIERS. Congressman Perkins wrote, "[I]s it realistic to try to serve the information needs of Federal (including the Congress), State, and local levels with information based on local agency evaluations?" (Note 12). Although OE officials had raised this question even prior to the enactment of the TIERS mandate, this inquiry by Perkins represented the Congress' first explicit acknowledgement of the problem. Congressman Perkins' letter echoed the language of a GAO report that would not be issued until September 1977. That report raised serious doubts about the ability of TIERS "to meet program information needs at Federal, local, and/or state levels" (p. 37). John Evans of OE responded to Perkins in August 1977 by acknowledging the difficulties inherent in the TIERS concept (Note 13). The Evans letter defended the agency's overall TIERS strategy, however, by making two points:

- OE intended that TIERS data on Title I be supplemented by national studies of Title I. While not generally relevant to state or local purposes, these studies were thought by Evans to be particularly useful in providing an accurate picture of the implementation and effectiveness of Title I activities nationwide.

- Evans stated his belief that the primary purpose of TIERS was "to upgrade State and local evaluation practices" and that the system was achieving that purpose. In that regard Evans said, "[T]he reaction to the models has in general been more positive than we expected. In many cases there has been acknowledgement of the superiority of the new methods over the old ways. In addition, use of the methods is increasing rapidly." He warned, however, that "when use of the models is required by program regulations and is no longer voluntary, there will probably be a more noticeably negative reaction."

The future of the TIERS requirements in Title I sparked little interest in hearings prior to reauthorization of ESEA, however (Subcommittee on Elementary, Secondary, and Vocational Education, 1978, pp. 118-176). Changes made in the national evaluation requirements affected mainly the sampling of students and the scheduling of assessments and reporting. These changes are indicated in Appendix A.

Thus, despite the questions raised in the GAO report and repeated by Congressman Perkins, congressional actions retained the TIERS requirements through the 1978 amendments with only minor modifications. After the 1978 reauthorization, congressional interest focused mainly on the development and timing of final regulations mandating local use of the TIERS models (Perkins, Quie, Ford, & Kildee, Note 14; Boyer, Note 15; Quie, Perkins, Ford, & Kildee, Note 16). As a result of congressional requests, OE disseminated the draft regulations for broad review prior to publishing them in the Federal Register in a Notice of Proposed Rulemaking on June 29, 1979 (44 FR 38400). Final regulations on Title I evaluation were published on October 12, 1979 (44 FR 59152) and published a second time on January 19, 1981 (46 FR 5160), along with final regulations for all other provisions of Title I. The final policy manual for Title I evaluation was published on December 22, 1980, and disseminated to all agencies affected by Title I evaluation requirements (Note 17).

In March 1980 the first of the three biennial reports to Congress on Title I evaluations required by the 1978 Title I amendments was transmitted in draft form to John F. Jennings, majority counsel of the House Subcommittee on Elementary, Secondary, and Vocational Education (Note 18).

In addition to general descriptive data on Title I, this report summarized Title I achievement data from 22 states. Though a month late, the draft report fulfilled the requirements in Section 183(g) that ED submit a report to the congressional committees responsible for Title I authorizations and appropriations by February 1, 1980. Because of the press of other responsibilities in ED, no effort was made to revise the report for issuance in final form to all members of the authorizations and appropriations committees (Anderson, Note 19). Although the tardiness of the formal evaluation report had been noted in a 1980 study of Title I federal administration (Reisner, Note 20), no Members of Congress or staff asked for copies of it (Anderson, Note 19).

Extent to which the mandate was implemented. An examination of the current Title I evaluation system indicates that the requirements for federal action in Section 183 have been met. 17/ To review briefly, the requirements pertaining to evaluation standards and schedules in Section 183(b), evaluation models in Section 183(d), and objective criteria in Section 183(f) were implemented through guidance provided in the final regulations for Title I evaluation. Specific requirements pertaining to these provisions were stated in the regulations and explained more fully in the evaluation policy manual for Title I. Requirements in Section 172(1) pertaining to state evaluation reports were also specified in the evaluation regulations and discussed in the policy manual. Requirements pertaining to local Title I evaluations, other than those specified in the evaluation regulations, were included in the Title I program regulations that were published in final in January 1981.

Requirements for federal provision of technical assistance in Title I evaluation were met through the creation of the TACs and through publication of monographs on various technical issues in the implementation of the evaluation requirements.

17/ The extent of state and local implementation of Section 183 presents a different set of questions, which are addressed in the preceding chapters of this volume.

Requirements for biennial reports to Congress on Title I evaluation have been generally met. As noted previously, the 1980 report was a month late, issued in draft form, and transmitted to only one congressional staff member. The 1982 report has been drafted and will be transmitted several months late but will meet other requirements of Section 182(g) (Stonehill, Note 21).

Requirements for the collection and dissemination of information on exemplary Title I projects, as stated in Section 183(h), have been met through a system that consists of project screening and dissemination procedures. Under these procedures, state Title I offices identify local Title I projects that they believe to be particularly effective in meeting Title I goals. Data on project effectiveness are then reviewed by the JDRP. If the project data indicate that the project is, in fact, exemplary, descriptive information on the project becomes part of the National Diffusion Network (NDN) and is disseminated broadly through the NDN system of SEA based "facilitators" and published periodicals. Although the JDRP and NDN systems are open to all elementary and secondary educational projects, Title I projects are given particular encouragement and prominence. In addition, the Title I program office in Washington has assigned a full-time liaison to encourage broad dissemination of the Title I projects approved by the JDRP.

Aspects of the 1974 mandate presenting the most serious problems. The most difficult problems in implementing TIERS arose from the dual purposes intended to be achieved by the system. (See also Wisler and Anderson, 1979, for a discussion of the tension inherent in the purposes of TIERS.) The requirement that TIERS data assist in local program improvement while also contributing to a national picture of Title I proved as difficult to implement as OE had predicted in 1975. The legislative requirements forced states and local school systems to change their assessment systems, sometimes radically, in order to meet certain objectives that were largely external to their own needs.

The technical design of the system was not what had made the implementation process difficult, even though the system had evolved in a gradual fashion as more was learned about techniques for project

assessment and data aggregation. According to federal observers, many SEAs and LEAs simply resisted implementing changes that were required to achieve national uniformity in the testing and analysis of the achievement of a large sector of elementary and secondary students. This resistence was overcome by a combination of two factors (Evans, Note 2):

- SEAs and LEAs realized that they were required by federal law to collect and report evaluation data according to TIERS requirements or else lose their Title I funds.
- At the same time they were given extensive assistance through the TACs in implementing the system and adapting it to meet their own constraints and needs.

Even though TIERS is now largely implemented, it is the belief of ED evaluation managers that the maintenance of the system will depend on the support for the system provided by ED (Stonehill, Note 21).

TIERS' Success as an Evaluation System

Whatever the success of the implementation of the mandate itself, the real test of the federal level effectiveness of this system is in its utility as a producer of evaluative information. To determine the effectiveness of TIERS in this regard, it is necessary to explore several different questions, as follows:

- What types of Title I information are needed in Congress and the Executive Branch?
- To what extent has TIERS provided such information?
- How, if at all, could TIERS be changed to increase its effectiveness in providing information to federal level users?

Before addressing these questions, it is important to state again that the Members of Congress who were responsible for drafting the 1974 evaluation mandate, who sponsored its passage, and who tracked its implementation by OE seem to have been as interested in the state and local uses of TIERS information as in its federal uses. This point is discussed more fully by Orland in his analysis of the local purposes motivating congressional endorsement of the provisions of Section 183 (Note 22). He argues that the 1978 amendments to the original TIERS

authorization emphasized the local uses and value of the evaluation system embodied in the legislative mandate. Although the discussion here looks only at federal uses of TIERS data, it is important to remember the importance placed on local uses by the congressional sponsors of the system.

Types of Title I information needed by Congress and the Executive Branch. In Congress the primary federal users of evaluative information on Title I are members and staff of the authorizations, appropriations, and budget committees. In the Executive Branch the primary users are ED managers of the program and of the overall departmental budget. Other Executive Branch users of Title I evaluations are the individuals in the Office of Management and Budget responsible for developing and defending ED budget requests. Information on the Title I and Chapter I data needs of all these individuals was collected through interviews and through the review of several broad examinations of evaluation usage at the federal level, most notably those conducted by Florio, Behrmann, and Goltz (1979), by Florio (1980, Note 23), by Boruch and Cordray (1980, Note 24), and by Raizen and Rossi (1981). From these sources it was possible to identify a fairly consistent picture of information needs. This inquiry indicated particular interest in information on the achievement impact of federally funded compensatory services. Some interest was also seen in the targeting of services on students and the cost of compensatory services. The achievement impact of Title I was, however, consistently described as the most important information need.

The centrality of student achievement as the key measure of program effectiveness was a somewhat surprising finding to this observer, for several reasons. First, the targeting of services has historically been a major concern under Title I. The 1978 debate over the targeting of Title I services on poor versus non-poor students and on low-achieving versus non-low-achieving students indicated the continuing relevance of the early view of Title I as a tool in the national War on Poverty (Subcommittee on Elementary, Secondary, and Vocational Education, pp. 170-175 and 184-189). From that perspective, the question of who receives Title I services is a primary concern in measuring the success of the program and must be addressed prior to the question of the effects of

the services. Second, because Chapter I changed the requirements for student and school targeting that had existed under Title I (Silverstein and McMullan, Note 25), it would be logical to expect keen interest in the effects of those changes on the actual selection of schools and students to receive federal compensatory services. Although interest was expressed in these questions by virtually all interview respondents, none of the respondents ranked student targeting as a more important question for evaluative purposes than student achievement.

This interest in student achievement under Title I conflicts in an interesting fashion with conclusions reached by Barnes and Ginsburg (1979) in their analysis of the relevance of TIERS for Title I policy concerns. Barnes and Ginsburg state:

> On its face the [TIERS] approach [to the collection and aggregation of local achievement data] appears to offer precisely what the ZBBS ["zero-based" budgeting system] advocates are calling for: a national measure of aggregate performance gains that are attributable to Title I projects. It is unlikely, however, that future appropriations levels will be influenced by these results. . . . The steady growth in appropriations over the years, despite very mixed evaluation results, is perhaps the best evidence of this (from $1.5 billion in 1965 to nearly $3 billion in 1979). Judging from past experience, negative findings are more likely to provoke criticisms of the evaluation effort than the program.

Today, the national picture is exactly opposite from that described by these analysts in 1979: Title I evaluation results are being interpreted as showing positive achievement gains for Title I participants, while appropriations levels are declining rapidly. According to interviews conducted for this study, this situation has prompted considerable interest in the overall achievement effects of the program--both now and over the next few years.

This interest was expressed succinctly by Richard DiEugenio, legislative aide to Representative John M. Ashbrook of Ohio, ranking minority member of the House Education and Labor Committee. DiEugenio stated that he expects the Committee to be particularly interested in how recent spending cuts in Chapter I affect the numbers of Chapter I participants and their achievement levels (Note 26). If the numbers of students being served and the achievement gains of these students change relatively little in the next several years, Congress will have to conclude that the

spending cuts did not hurt students. If, however, it becomes clear that substantially fewer students are being served and that the achievement gains of participants are smaller or even nonexistent, Congress may want to question the wisdom of the spending cuts. (Although not noted by DiEugenio, Chapter I effects may be a factor as well in future assessments of the provisions of Chapter I itself.)

Congressional interest in student achievement in Title I and Chapter I is consistent with findings of Florio, Behrmann and Goltz. In their survey of congressional staff members who work with education legislation, these researchers found that the type of program data valued most highly by their respondents was information on program impact. Moreover, within that category "the most significant information sought by staff is the impact on student achievement, particularly in the basic skills" (p. 76).

In ED the need for Chapter I data is consistent with the patterns seen in Congress. ED program managers and their staff express continuing interests in the student achievement impact of federal compensatory services. In the ED Office of Compensatory Education Programs, data on state-wide achievement levels as reported through TIERS are provided to state Title I coordinators and to federal program officers responsible for overseeing Title I expenditures in each state (Lobosco & Enderlein, Note 27). Information on student achievement has been made available primarily to offer a means for each state to compare its own Title I performance with overall national achievement performance under Title I.

The ED budget office expressed somewhat greater interest in student targeting as an evaluation question than did other offices (Brown, S., & Brown, L., Note 28). This interest is related to that office's responsibility to estimate patterns of service coverage at varying funding levels for Chapter I. Interview respondents noted, however, that achievement oriented questions are more frequently asked by members of the congressional appropriations committees than are other types of questions.

Information provided by TIERS to date. As discussed previously, TIERS is only now generating state Title I data that can be aggregated to provide a comprehensive picture of Title I operations and effectiveness nationwide. The FY 1981 Annual Evaluation Report on Education Programs utilizes data from TIERS in reporting on the following five topics:

- Types of services (e.g., reading, mathematics, support services, etc.) received by Title I participants;
- Parent activities related to the local Title I projects, including activities of local Title I parent advisory councils;
- Staff training supported by Title I;
- Reading achievement of Title I students in grades two through eight; and
- Mathematics achievement of Title I students in grades two through eight.

In the FY 1981 Annual Evaluation Report, ED promises to provide future analyses of the following topics using TIERS data:

- Reading achievement of Title I students in grades 9-12;
- Mathematics achievement of Title I students in grades 9-12;
- Language arts achievement (other than reading) of Title I students in grades 1-12;
- Analyses of the relationship of achievement gains to summer school participation; and
- Analyses of program characteristics of reading and mathematics projects in grades 2, 6, and 10, including—
  - instructional strategies,
  - amount of time devoted to Title I-funded instruction, and
  - student-teacher ratios. 18/

These categories of information appear to conform fairly well with the stated needs of congressional and ED managers and staff having Title I responsibilities. Achievement results of Title I are addressed in at least a preliminary fashion by this design. Important descriptive information on program activities will also be generated.

Other types of Title I information are not at present being collected and reported by TIERS, however, such as the following:

- Longitudinal information on the cumulative effects of Title I participation over a multi-year period;
- Information that would permit achievement effects to be determined according to variations in--

---

18/ These analyses are expected to be presented in the forthcoming ED report to Congress on Title I.

- instructional strategies,
- amounts of daily or weekly time on reading or mathematics tasks,
- student teacher ratios, and
- teacher characteristics;

. Information on the family background of students; and

. Information on the cost of services.

Data on topics such as these are being generated by the large national study of the sustaining effects of compensatory education. Because of the more highly controlled nature of their data collection procedures, the sustaining effects data will permit the development of a much broader range of evaluative insights into Title I program effects than is currently possible with data generated by TIERS. On the other hand a national sample design such as that used for the sustaining effects study cannot produce data that are valid on a state-by-state basis. These and other considerations must be balanced to determine the extent to which national evaluation needs for Title I require greater emphasis on a locally based approach such as TIERS or on a national sample study such as the sustaining effects study.

Approaches to increasing the utility of TIERS for federal users of evaluation data. A key question in upcoming policy debates over Chapter I will be the character and extent of program change resulting from the new requirements and the decreased levels of program funding. The school year 1979-80 TIERS data reported recently by ED will provide important baseline information for future comparisons. To be most valuable, future data should permit conclusions to be drawn regarding the following:

. Changes in the types of services rendered;

. Changes in the intensity of services (e.g., amount of daily/weekly instructional time and student-teacher ratios);

. Changes in the characteristics of instructional staff (e.g., shift to less experienced, more poorly trained staff);

. Changes in numbers of students served; and

. Changes in background characteristics of students served (e.g., family income, initial level of educational disadvantagement, etc.).

The continuation in present form of the TIERS data collection and analysis activities would permit all but the last of these categories of information to be generated.

TIERS as a Model for Other Program Evaluation and Reporting Systems

Federal experience with TIERS has had some relevance to the design and operation of evaluation and reporting systems in other education programs. This section of the chapter discusses the extent to which TIERS has acted as a model in this regard. It also discusses the ways in which the federal experience with TIERS is relevant to grantee evaluation and reporting systems that might be established under new administrative structures for education at the federal level, including program consolidation and block grants.

TIERS as a model for current evaluation and reporting systems in other federal education programs. No other federal education program requires systematic evaluation and reporting by local grantees on a scale similar to TIERS. This conclusion is apparent from an examination of federal evaluation and reporting requirements imposed on state and local administrative agencies. Two factors are particularly important in this regard. First, only a handful of programs are large enough in total funding level and numbers of local grantees to qualify as suitable comparisons to Title I. These programs include those providing aid for education of the handicapped, aid for vocational education, and aid to federally impacted areas. Second, none of these programs generates as uniform a set of local services as Title I, with its broad and consistent emphasis on compensatory instruction in basic skills. Thus, none of the other large programs has as uniform a set of local services to compare from one site to the next as does Title I.

Several smaller programs do, however, offer interesting programmatic analogies to TIERS, in the sense that, first, all their local grantees offer relatively similar services; second, there are consistent similarities among all students served; and third, reporting and evaluation requirements are explicitly imposed by the programs. These programs are those offering aid for bilingual education, for the education of Indian students, and for basic skills improvement. Although there are program

components similar to TIERS in each of these programs, none requires that local school systems generate effectiveness data that can be aggregated at state or federal levels. Because of that central omission, none of these programs may be said to have adopted the TIERS model, though all have been affected by it. For example:

- Under the Bilingual Education Act (ESEA Title VII), models for the evaluation of local program effectiveness have been developed and disseminated. According to the federal director of that effort, the Title I evaluation models were used extensively in the development of the Title VII models (Shoemaker, Note 29). As noted previously, however, the bilingual program did not structure its evaluation design to permit the national aggregation of local evaluation results, thus omitting an essential feature of the Title I evaluation framework. In addition to the use of evaluation models, a second similarity to Title I was the establishment of technical assistance centers, known as Evaluation, Dissemination, and Assessment Centers, to provide aid in the design and conduct of local evaluation activities.
- Technical assistance for evaluation purposes is also a feature of the evaluation activities conducted under the Indian Education Act. As in Titles I and VII, technical assistance centers have been established, in part, to provide aid for the evaluation of local activities funded under the Indian Education Act. These centers, which are five in number, provide technical assistance to LEAs on a number of different topics, concentrating mainly on general program improvement (Shuler, Note 30). No efforts have been made to develop uniform approaches to local evaluation, however, as was the case in Title I. At the time that these Indian Education Centers were established in 1979, consideration was given to merging the functions of the centers into the Title I TACs (Beebe, Note 31). Because evaluation was only one of several program improvement priorities to be met, however, the decision was made to establish a separate set of Indian centers.

TIERS as a model for evaluation and reporting systems that might be part of current and future program consolidation structures and block grants. In the past year one major program consolidation has been achieved in education. ECIA Chapter II consolidates 29 education programs in a framework that significantly increases state and local discretion over the expenditure of federal funds. Other program consolidations have been proposed in connection with vocational and adult education and education of the handicapped. If previous federal experience with consolidations and block grants is a guide to the future, two forces are likely to create noticeable tensions in the design and implementation of any reporting or evaluation requirements in these new assistance structures.

The first such force is the consistent objective of the designers of block grants and consolidations to reduce red tape for the local recipients of federal aid (Stenberg, 1977, p 6). This objective is cited in Section 561(a) of ECIA Chapter II:

> It is the purpose of this chapter to consolidate the program authorizations contained in [programs listed]. . . and to do so in a manner designed to greatly reduce the enormous administrative and paperwork burden imposed on schools at the expense of their ability to educate children.

The second force is described by Stenberg as "creeping categorization" (p. 41):

> . . . [A]s block grants age, they become subject to "creeping categorization," as the Congress, interest groups, and federal administrators seek to emphasize specific national priorities, guarantee funding of popular programs, or raise appropriations levels. . . . This phenomenon appears to be a significant factor in understanding the "life cycle" of such grants and makes for a rather unstable instrument.

The effect of "creeping categorization" on grantee paperwork requirements is addressed by Hastings (Note 34) in her review of the literature on block grants:

> While in some block grants paperwork seems to have been reduced, in others, such as the community development block grant, many administrators reported no change or an actual increase in paperwork. . . . Moreover, as the block grants have come up for reauthorization, Congress frequently has added new controls to restrict the actual or alleged misuse of federal funds. . . . These new controls often have added paperwork.

Thus, one of the key objectives of the block grant mechanism--paperwork reduction--appears to be particularly difficult to achieve.

This phenomenon suggests that grantee reporting and evaluation requirements are likely to be drastically reduced or even eliminated as categorical programs are transformed into block grants. However, when categorical interests reassert themselves in congressional appropriations decisions, in oversight by congressional authorizing committees, and in other ways, grantee evaluation and reporting are likely to be imposed, if only to yield visibility to the special purposes and groups whose needs have been communicated to Congress and the Executive Branch. Stenberg suggests the benefits that such requirements can achieve in a block grant or program consolidation context (p. 39):

> . . . [W]here congressional committees require the federal administering agency to evaluate and report periodically on recipient performance, opportunities are provided to adjust funding policies and priorities to better reflect national objectives via administrative rather than statutory action.

Information on which to base such evaluation would have to come from local projects and would require specific recordkeeping and reporting standards for local administrators. Thus, although evaluation and reporting requirements are not certain to be reimposed in future education block grants, there is no persuasive reason to believe they won't.

Future Federal Needs for TIERS

The transition from Title I to Chapter I is not likely to reduce congressional and Executive Branch needs for information on the characteristics and effects of federally-supported compensatory education services. As stated by one congressional aide, certain assumptions now prevail in Congress that the system for generating Title I (or Chapter I) information is in place and working (DiEugenio, Note 26). If anything, federal users of that information will have even greater needs for those data in the future than in the past, as they examine the effects of the transition to Chapter I and the reductions in federal compensatory funding.

This is not to say that TIERS should remain a static system, however. As discussed elsewhere in this report, changes can usefully be made in many aspects of the data collection and data analysis components of TIERS and the associated system of technical assistance. New consideration will

have to be given to the relationship between the roles played by TIERS and by the national studies of federal compensatory education, especially the suitability of each for addressing specific evaluative questions. The overall system embodied in TIERS is now, however, an essential aspect of the federal structure for compensatory education support. If it were fully dismantled today, it is exceedingly probable that it would soon have to be created all over again.

REFERENCE NOTES FOR CHAPTER SIX

1. Jennings, J. F. Personal communication, January, 1982.

2. Evans, J. W. Personal communication, February, 1982.

3. Ottina, J. Testimony before the House Education and Labor Committee, May 22, 1974.

4. U. S. Office of Education. Progress report on the implementation of Title I evaluation requirements of the new Section 151 of the Elementary and Secondary Education Act as amended by P. L. 93-380. Unpublished report, January 31, 1975.

5. Cross, C. T. Letter to T. H. Bell, Commissioner of the U. S. Office of Education, Washington, D.C., February 6, 1975.

6. Evans, J. W. Letter to legislative aide C. T. Cross, House of Representatives, Committee on Education and Labor, Washington, D. C., March 20, 1975.

7. Perkins, C. D., & Quie, A. H. Letter to T. H. Bell, Commissioner of the U. S. Office of Education, October 16, 1975.

8. Bell, T. H. Letter to A. H. Quie, House of Representatives, Committee on Education and Labor, February 24, 1976.

9. U. S. Office of Education. The development of state and local Title I evaluation models and reporting systems. ESEA Title I program directive, June 23, 1976.

10. Lind, M. L., Colburg, D., Clark, C. G., Duncan, V. A., Truby, R., & Brouillet, F. B. Letter to R. Wheeler, Deputy Commissioner for the Bureau of School Systems, U. S. Office of Education, September 8, 1976.

11. Nyquist, E. B. Letter to W. Pierce, Acting Commissioner of the U. S. Office of Education, September 21, 1976.

12. Perkins, C.D. Letter to J. W. Evans, Director of Evaluation, U. S. Office of Education, July 21, 1977.

13. Evans, J. W. Letter to C. D. Perkins, House of Representatives, Committee on Education and Labor, August, 1977.

14. Perkins, C. D., Quie, A. H., Ford, W. D., & Kildee, D. E. Letter to E. L. Boyer, Commissioner of the U.S. Office of Education, May 23, 1978.

15. Boyer, E. L. Letter to C. D. Perkins, House of Representatives, Committee on Education and Labor, June 14, 1978.

16. Quie, A. H., Perkins, C. D., Ford, W. D., & Kildee, D. E. Letter to E. L. Boyer, Commissioner of the U. S. Office of Education, June 22, 1978.

17. RMC Research Corporation. Evaluation policy manual for ESEA Title I. Prepared for the U. S. Department of Education. Mountain View, Cal., December 22, 1980.

18. U. S. Office of Education, Office of Evaluation and Dissemination. Title I of the Elementary and Secondary Education Act: Evaluation report to Congress as required by Section 183 of ESEA Title I. Unpublished report in draft, March 1980.

19. Anderson, J. K. Personal communication, January 1982.

20. Reisner, E. R. The Office of Education administers changes in a law: Agency response to the Education Amendments of 1978. Prepared for the U. S. Department of Education, National Advisory Council on the Education of Disadvantaged Children. Washington, D. C.: NTS Research Corporation, May, 1980.

21. Stonehill, R. M. Personal communication, January 1982.

22. Orland, M. E. Evaluating ESEA Title I: Local responses to federal expectations. Unpublished manuscript, July 1979. (Available from NTS Research Corporation, Durham, N.C.).

23. Florio, D. H. Educational inquiry and the Congress: Unmet potential or undervalued resources? In Educational inquiry and the legislative process. Symposium presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

24. Boruch, R. F., & Cordray, D. S. An appraisal of educational program evaluations: Federal, state and local agencies. Prepared for the U. S. Department of Education. Evanston, Ill.: Northwestern University, June 30, 1980.

25. Silverstein, R., & McMullan, S. The Education Consolidation and Improvement Act of 1981: Its meaning for state and local policymakers and administrators. Prepared for the National Association of State Boards of Education. Washington, D. C., 1981.

26. DiEugenio, R. Personal communication, January 1982.

27. Lobosco, W., & Enderlein, T. Personal communication, January 1982.

28. Brown, S., & Brown, L. Personal communication, January 1982.

29. Shoemaker, D. Personal communication, January 1982.

30. Shuler, D. Personal communication, February 1982.

31. Beebe, C. Memorandum to the Commissioner of the U. S. Office of Education, March 30, 1979.

32. O'Leary, H. Personal communication, January 1982.

33. Wohlstetter, P., & Boruch, R. The origins, intensity, and impact of congressional oversight: The case of Title I. Unpublished manuscript in draft, November 1981. (Available from Northwestern University, School of Education.)

34. Hastings, A. H. The block grant record: Expectations, experience, and implications. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April 1981.

REFERENCES

Alderman, D. L., Swinton, S. S., & Braswell, J. S. Assessing basic arithmetic skills and understanding across curricula: Computer-assisted instruction and compensatory education. The Journal of Children's Mathematical Behavior, 1979, 2, 3-28.

Anderson, J. K., Johnson, R. T., Fishbein, R. L., Stonehill, R. M., & Burns, J. C. The U. S. Office of Education models to evaluate ESEA Title I: Experiences after one year of use. Washington, D.C.: U. S. Office of Education, 1978.

Alkin, M. C., Daillak, R., & White, P. Using evaluations: Does evaluation make a difference? Beverly Hills, Cal.: Sage Publications, 1979.

Baglin, R. F. Does "nationally" normed really mean nationally? Journal of Educational Measurement, 1982, 18, 97-108.

Barnes, R. E., & Ginsburg, A. L. Relevance of the RMC models for Title I policy concerns. Educational Evaluation and Policy Analysis, 1979, 1(2), 7-14.

Boruch, R.F., Cordray, D.S., Pion, G., & Leviton, L. A mandated appraisal of evaluation practices: Digest of recommendations to the Congress and the Department of Education. Educational Researcher, 1981, 10, 10-13, 31.

Committee on Education and Labor, U. S. House of Representatives. Elementary and Secondary Education Amendments of 1974. House of Representatives document 93-805. 93rd Congress, 2nd Session. Washington, D. C.: U. S. Government Printing Office, February 21, 1974.

Committee on Education and Labor, U. S. House of Representatives. A Report on the Education Amendments of 1978, H. R. 15. House of Representatives document 95-1137. 95th Congress, 2nd Session. Washington, D. C.: U. S. Government Printing Office, May 11, 1978.

Committee on Human Resources, U. S. Senate. Education Amendments of 1978: Report to Accompany S. 1753. Senate document 95-856. 95th Congress, 2nd Session. Washington, D. C.: U. S. Government Printing Office, May 15, 1978.

Cross, C. T. Title I evaluation - A case study in congressional frustration. Educational Evaluation and Policy Analysis, 1979, 1(2), 15-21.

Cronbach, L. J. Course improvement through evaluation. Teachers College Record, 1963, 64, 672-683.

Florio, D. H., Behrmann, M. M., & Goltz, D. L. What do policy makers think of educational research and evaluation? Or do they? Education Evaluation Policy Analysis Journal of the American Educational Research Association, Nov.-Dec. 1979, 1(6), 61-87.

Hoepfner, R. Achievement test selection for program evaluation. In M. J. Wargo & D. R. Green (Eds.) Achievement testing of disadvantaged and minority students for educational program evaluation. Monterrey, Cal.: CTB/McGraw-Hill, 1978.

House, E. H. A review of reports from the National Institute of Education on the Compensatory Education Study (6 Volumes). (A review of Proceedings of the National Academy of Education (Vol. 6).). Washington, D.C.: National Academy of Education, 1979.

Leinhart, G., & Seewald, A. M. Overlap: What's tested, what's taught? Journal of Educational Measurement, 1981, 18, 85-96.

Linn, R. L. Discussion: Regression toward the mean and the interval between test administrations. In G. Echternacht (Ed.), New Direction for Testing and Measurement, 1980, 8, 83-89.

McLaughlin, M. W. Evaluation and reform. Cambridge, Mass.: Ballinger Publishing Co., 1975.

Millman, J., Paisley, W., Rogers, W. T., Sanders, J. R., & Womer, F. B. Performance review of USOE's ESEA Title I evaluation technical assistance program. Washington, D.C.: Department of Health, Education, and Welfare, April 1979.

National Center for Education Statistics (NCES/FRSS). School district needs for technical assistance. Washington, D.C.: U. S. Department of Health, Education and Welfare, 1980 (NCES-80-231).

National Institute of Education. Compensatory education services. Washington, D.C.: NIE, 1977.

Raizen, S., & Rossi, P. (Eds.). Program evaluation in education: When? how? to what ends? Washington, D. C.: National Academy of Sciences, 1981.

Roberts, A. O. H. Regression toward the mean and the regression-effect bias. In G. Echternacht (Ed.), New Directions in Testing and Measurement. 1980, 8, 59-82.

Rosenthal, R. Experimenter Effects in Behavioral Research. New York: Appleton, Century Crofts, 1966.

Rost, P. Using evaluation data for program planning in Title I. Prepost Press, 1980, V(2), 5.

Stenberg, C. Block grants: A comparative analysis. Prepared for the Advisory Commission on Intergovernmental Relations. Washington, D. C.: U. S. Government Printing Office, October 1977 (ACIR Publication No. A-60.)

Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor; U. S. House of Representatives. Part 16 of hearings on H. R. 15: Services and student development in ESEA Title I. 95th Congress, 1st Session. Washington, D.C.: U.S. Government Printing Office, 1978.

Tallmadge, G. K., & Wood, C. T. User's Guide: ESEA Title I Evaluation and Reporting System. Mountain View, Cal.: RMC Research Corporation, October, 1976.

Tallmadge, G. K., & Wood, C. T. User's Guide: ESEA Title I Evaluation and Reporting System. Mountain View, Cal.: RMC Research Corporation, 1981.

U.S. Comptroller General, General Accounting Office. Problems and needed improvements in evaluating Office of Education programs. Washington, D. C.: U. S. Government Printing Office, September 8, 1977 (HRD-76-165).

U.S. Department of Education. FY 1981 annual evaluation report on education programs. Washington, D.C.: Author, November 1981.

Wang, M. M. The nature of recipients of compensatory education. (Report No. 3 of the Study of the Sustaining Effects of Compensatory Education on Basic Skills.) Washington, D.C.: U. S. Department of Education, 1978.

Wisler, C. E., & Anderson, J. K. Designing a Title I evaluation system to meet legislative requirements. Educational Evaluation and Policy Analysis, 1979, 1(2), 47-55.

GLOSSARY OF ABBREVIATIONS

ECIA	Education Consolidation and Improvement Act
ED	U. S. Department of Education
ESEA	Elementary and Secondary Education Act
FTE	full-time equivalent
FY	fiscal year
GAO	U. S. Government Accounting Office
HEW	U. S. Department of Health, Education, and Welfare
JDRP	Joint Dissemination Review Panel
LEA	local educational agency
NCE	Normal Curve Equivalent
NCES	National Center for Education Statistics
NDN	National Diffusion Network
NIE	National Institute of Education
OE	U. S. Office of Education
SEA	state education agency
TAC	Technical Assistance Center
TIERS	Title I Evaluation and Reporting System
ZBBS	"zero based" budgeting system

APPENDIX A

Title I Evaluation Requirements, As Contained in the Education Amendments of 1978 [1/]

Section 124(g) Evaluations

A local educational agency may receive funds under this title only if—

(1) effective procedures are adopted for evaluating, [in accordance with the evaluation schedule promulgated by the Commissioner under section 183(g),] the effectiveness of the programs assisted under this title in meeting the special educational needs of educationally deprived children;

[(2) such evaluations will include, during each three-year period, the collection and analysis of data relating to the degree to which programs assisted under this title have achieved their goals, including the requirements of section 130 [participation of children enrolled in private schools], and will also include objective measurements of educational achievement in basic skills over at least a twelve-month period in order to determine whether regular school year programs have sustained effects over the summer; and

(3) the evaluation will address the purposes of the programs, including the requirements of section 130, and the results of the evaluations will be utilized in planning for and improving projects and activities carried out under this title in subsequent years.]

Section 172(1) Reporting

Each State educational agency shall make to the Commissioner periodic reports (including the results of objective measurements required by section 124(g) and of research and replication studies) evaluating the effectiveness of payments under this title and of particular programs assisted under it in improving the educational attainment of educationally deprived children.

Section 183 Program Evaluation

(a) INDEPENDENT EVALUATIONS.-- The Commissioner shall provide for independent evaluations which describe and measure the impact of programs and projects assisted under this title. Such evaluations may be provided by contract or other arrangements, and all such evaluations

---

1/ Text appearing in heavy brackets was added to Title I by the 1978 amendments (P.L. 95-561). Other text in Section 124(g) and Section 172(1) had been present in the law prior to 1974. Text in Section 183 not appearing in heavy brackets had been added to Title I by the 1974 amendments (P.L. 93-380).

shall be made by competent and independent persons, and shall include, whenever possible, opinions obtained from program or project participants about the strengths and weaknesses of such programs and projects.

(b) EVALUATION STANDARDS AND SCHEDULE.-- The Commissioner shall (1) develop and publish standards for evaluation of program or project effectiveness in achieving the objectives of this title, [and (2) develop, in consultation with State educational agencies and representatives of local educational agencies, a schedule for conducting evaluations under section 124(g) designed to ensure that evaluations are conducted in representative samples of the local educational agencies in any State each year. Such standards will be developed only after widespread consultation and hearings with practicing State and local agency evaluators, and the Commisisoner's standards will reflect the input of these groups.]

(c) JOINTLY SPONSORED STUDIES.-- The Commissioner shall consult with State [and local] educational agencies in order to provide for jointly sponsored objective evaluation studies of programs and projects assisted under this title within a State.

(d) EVALUATION MODELS.-- The Commissioner shall provide to State educational agencies, models for evaluations of all programs conducted under this title, for their use in carrying out their functions under section 172 [requirements for State reports to the Commissioner], which shall include uniform procedures and criteria to be utilized by local educational agencies and State agencies as well as by the State educational agency in the evaluation of such programs. [In developing evaluation design models the Commissioner shall consult with State and local evaluators experienced in conducting such evaluations.]

(e) TECHNICAL ASSISTANCE.-- The Commissioner shall provide such technical and other assistance as may be necessary to State educational agencies to enable them to assist local educational agencies [and State agencies] in the development and application of a systematic evaluation of programs in accordance with the models developed by the Commissioner.

(f) SPECIFICATION OF OBJECTIVE CRITERIA.-- The models developed by the Commissioner shall specify objective criteria which shall be utilized in the evaluation of all programs and shall outline techniques (such as longitudinal studies of children involved in such programs) and methodology (such as the use of tests which yield comparable results) for producing data which are comparable on a statewide and nationwide basis.

(g) REPORT TO CONGRESS.-- The Commissioner shall make a report to the respective committees of the Congress having legislative jurisdiction over programs authorized by this title and the respective Committees on Appropriations no later than [February 1, 1980, 1982, and 1984] concerning the results of evaluations of programs and projects required under this section, which shall be comprehensive and detailed, as up-to-date as possible, and based to the maximum extent possible on objective measurements, together with other related findings and evaluations and his recommendations with respect to legislation.

(h) INFORMATION DISSEMINATION.-- The Commissioner shall also develop a system for the gathering and dissemination of the results of evaluations and for the identification of exemplary programs and projects, or of particularly effective elements of programs and projects, and for the dissemination of information concerning such programs and projects or such elements thereof to State agencies and local educational agencies responsible for the design and conduct of programs and projects under this title, and to the education profession and the general public.

(i) MAXIMUM EXPENDITURES.-- The Commissioner is authorized, out of funds appropriated to carry out this title in any fiscal year, to expend such sums as may be necessary to carry out the provisions of this section, but not to exceed one-half of 1 per centum of the amount appropriated for such programs. [In carrying out the provisions of this section, the Commissioner shall place priority on assisting States, local educational agencies, and State agencies to conduct evaluations and shall, only as funds are available after fulfilling that purpose, seek to conduct any national evaluations of the program.] 2/

2/ This sentence was added in an amendment enacted in 1976 (P. L. 94-273).

APPENDIX B

Chronology of Communications

Between OE and Congress on TIERS [1/]

[1971	Contract awarded to American Institutes of Research (AIR) to analyze state Title I reports.]
[1972	AIR reports no evidence of Title I success on basis of state reports. AIR says that current state reports do not permit national conclusions to be drawn due to poor quality and lack of comparability of the reports.]
5/10/74	Letter from Congressman Carl D. Perkins of House Education and Labor Committee to OE Commissioner John Ottina. Requests testimony on evaluation issues, including the Veysey amendment regarding Title I evaluation.
5/17/74	Testimony of Ottina before House Education and Labor Committee on evaluation provisions of H.R. 69. Describes plans for implementing Veysey requirements but notes problems in national aggregation of locally collected data.
5/22/74	Letter from Perkins to HEW Secretary Casper Weinberger. Encloses 16 questions on evaluation to be answered by HEW representatives, including a question asking for an interpretation of the proposed Veyzey amendment.
No date	HEW's response to Perkins' letter is available only in draft form. However, a 12/13/77 letter from Janice Anderson of the OE evaluation office to Richard Jaeger of the University of North Carolina summarizes and cites relevant portions of the HEW letter. The quoted material from that correspondence is consistent with the undated draft.
[6/74	RMC awarded a contract to develop a standard state reporting form for Title I, which is to include models for the reporting of evaluation results.]

1/ Important dates and events in the development and implementation of TIERS are noted in brackets.

8/24/74 Enactment of Education Amendments of 1974 (P.L. 93-380), which include the TIERS mandate in Section 151.]

[1975 RMC completes its study and presents three results of its work:

- A standard report form for descriptive and impact data on Title I
- Three models for the reporting of impact data
- A set of suggested state requirements regarding test selection procedures, timing, metric, analysis and interpretation

RMC is awarded a follow-on contract to present the recommended procedures to all SEAs and to three LEAs in each state.]

1/31/75 Issuance of status report on the implementation of Section 151.

2/6/75 Letter from Christopher Cross, staff member to Congressman Albert Quie of the House Education and Labor Committee, to OE Commissioner Terrell Bell.

Concerned that Bell not emphasizing the mandatory use of the evaluation models. Wants to know schedule for models. Concerned with overlap between Sustaining Effects Study and NIE study of compensatory education.

3/20/75 Letter from OE evaluation director John Evans to Cross.

Evans believes that OE can't mandate and get aggregatable data from LEAs. Prefers national sample evaluations. However, plans to visit every state and three LEAs in each state to determine:

- Could they implement models?
- When?
- How much would it cost?

5/8/75 Memo from Bell to Chief State School Officers.

Describes plans for the implementation of the evaluation requirements contained in Section 151. Encloses a copy of the January status report on Section 151.

No date Letter from Congressman Daniel Flood of the House Appropriations Committee to OE requesting information on Title I studies and evaluations (not available).

6/27/75 Response to questions from Congressman Flood on Title I studies, listing all current FY 76 studies, their completion dates, and their costs.

10/16/75 Letter from Perkins and Quie to Bell expressing concern over current OE priorities for the evaluation of Title I.

Want a reduction of size and cost of Sustaining Effects study and greater attention to technical assistance to SEAs and LEAs.

10/24/75 Letter from Quie to Bell.

Reaffirms support for Title I reading and math, thus clarifying 10/16/75 letter.

2/24/76 Letters from Bell to Perkins and Quie.

In response to letter of 10/16/75, describes plans for a reduction in Sustaining Effects Study and increase in technical assistance for Title I evaluation. Encloses table indicating allocation of Section 151 funds.

[6/23/76 Issuance of OE Title I Program Directive.

Describes status of implementation of Section 151 and announces workshops on Title I evaluation models to be held the fall of 1976.]

8/12/76 Letter from Evans to Perkins and Quie.

Transmits the status report on Title I evaluation models dated 7/26/76.

9/8/76 Letter of complaint to OE from several Chief State School Officers:

- State opposition to adoption of standarized evaluation models
- Question technical soundness of models and adequacy of method of development
- Want OE to delay implementation of new evaluation requirements

9/21/76 Letter from Ewald Nyquist, New York Commissioner of Education, to OE.

Opposes the imposition of the proposed regulations. Recommends that final decision on pending regulations be delayed until evidence of need is determined.

[Fall 1976 TAC operations begin.]

1/10/77 Letter from Evans to Perkins and Quie.

Describes progress in implementation of Section 151:

- Established 10 regional TACs
- Increased public awareness through
  - TACs
  - workshops
  - handbooks on practical evaluation issues
  - newsletter
- Regulations being prepared; regs process to include
  - Notice of Intent to Regulate
  - Notice of Proposed Rulemaking by June 1977
  - Final regulations by late fall 1977
- Decision of 20 states to use models
- Increased funding for technical assistance
- Believes that regs "may generate considerable negative reaction"

Attaches tables of proposed funding patterns for various Section 151 activities.

7/21/77 Letter from Perkins to Evans asking six questions on planning and evaluation in education (not available).

7/29/77 Letter from Evans to Quie.

Transmits information requested by Cross regarding progress of implementation of Title I evaluation requirements:

- Three evaluation models developed, examined, and explained to educators
- Use of models expanded to sites in all states
- Implementation aided by OE through handbooks, workshops, and consulting services from TACs
- TACs and states have defined relationship; TACs have given 200 workshops
- Study of Sustaining Effects reduced
- Notice of Intent planned for publication in the fall

Encloses report by OE on the implementation of Section 151 and facts and figures regarding Section 151 which were placed as an insert into the transcript of the oversight hearings.

8/5/77 Letter from Evans to Perkins.

Transmits answers to questions from Perkins (7/21/77, reported above) on planning and evaluation in education. Articulates the advantages of the models (local use

for program improvement) and the disadvantages (problems of aggregation at the state and national levels).

10/7/77 Letter from Perkins to OE legislative director Al Alford.

Request for OE representatives to testify at hearings:

- Representative from Office of Planning, Budget and Evaluation re evaluation of Title I services and student development including those conducted under Section 151
- Representatives from Title I program administration re instructional issues in Title I
- Panel of OE Title I program people re administration of Title I

10/18/77 Testimony of Evans before Perkins.

Summarizes status of Title I evaluation activities; estimates that 40 percent of LEAs are using models.

4/28/78 Memo from Evans to Chief State School Officers

Requests comments on four issues concerning the implementation of Section 151 and describes early results of participation study. Four issues are:

- How models should be described
- Allowable variations and alternatives
- Use of Title I funds for testing students not participating in Title I
- Change in deadline for SEA Title I reports

States that a Notice of Intent will not be issued. Will complete draft regs by 6/5/78.

5/23/78 Letter from Perkins to OE Commissioner Ernest Boyer.

- Requests Committee review of regulations before publication in Federal Register
- Inquires as to plans for use and dissemination of evaluation data
- Asks if existing evaluation systems will have to be modified or abandoned in order to accommodate proposals of OE
- Requests OE plans for publication of evaluation regulations in addition to those mentioned in 4/28/78 memo

6/14/78 Letter from Boyer to Perkins responding to 5/23/78 request.

- Will disseminate draft regulations and sponsor a meeting to discuss models before publication in Federal Register
- Evaluation data will be used
  - to supply Congress with an annual report required by Section 151
  - to produce a "Title I Digest"
  - to identify effective projects
- States and districts will have to make slight modifications to their existing practices
- Dissemination of draft regulations prior to publication will result in a delay in the implementation process

6/22/78 Letter from Quie and Perkins to Boyer.

Request that final regulations be published no later than March 1979.

7/21/78 Memo from Evans to Chief State School Officers with draft of TIERS regulations enclosed.

- Asks for comments and suggestions regarding draft of proposed regulations
- Draft describes major requirements of the Section 151 evaluation and reporting system
- Technical Manual providing guidance regarding appropriate ways of implementing regulations to be disseminated late the following fall

8/14/78 Memo from Evans to Chief State School Officers.

- Requests reactions to draft regulations
- Meeting to be held in September with representatives of Committee of Title I State Coordinators and Committee on Evaluation and Information Systems to discuss technical and procedural aspects of their work
- Encloses report on experiences of state and local administrators in using Title I evaluation models during 1976-77 school year, prepared by OE staff

[11/1/78 Enactment of Education Amendments of 1978 (P.L. 95-561), which changes Section 151 to Section 183 and adds a new sampling provision and other changes.]

[1978 Draft regulations for Title I evaluation circulated. Preamble states that about 55 percent of all Title I LEAs are trying the models.]

[February 1979 Notice of Proposed Rulemaking for Title I evaluation issued in the Federal Register, LEA models and reporting system to become operational in fall 1979. OE sends sampling guidance to states; requires state sampling plans in April.]

[Spring 1979 Panel Report of the HEW Assistant Secretary for Planning and Evaluation on technical assistance, especially TACs, is issued (Millman et al, 1979).]

[10/12/79 Regulations for Title I evaluation published in final.]

[1980 Representative samples of LEAs complete first year of data collection and report to SEAs.]

[March 1980 Report (in draft) submitted to certain congressional offices as required by Section 183. Required to have been submitted 2/1/80. Reports on TIERS data from some states.]

[December 1981 ED issues FY 1981 Annual Evaluation Report on Education Programs, in which TIERS data from all 50 states are reported for the first time.]

APPENDIX C

Evaluation Requirements, As Contained in

the Education Consolidation and Improvement Act [1/]

CHAPTER I:

Section 555(d) Records and Information

Each State educational agency shall keep such records and provide such information to the Secretary as may be required for fiscal audit and program evaluation (consistent with the responsibilities of the Secretary under this chapter).

Section 556(b) Application Assurances

The application described in subsection (a) shall be approved if it provides assurances satisfactory to the State educational agency that the local educational agency will keep such records and provide such information to the State educational agency as may be required for fiscal audit and program evaluation (consistent with the responsibilities of the State agency under this chapter), and that the programs and projects described

. . .

(4) will be evaluated in terms of their effectiveness in achieving the goals set for them, and that such evaluations shall include objective measurements of educational achievement in basic skills and a determination of whether improved performance is sustained over a period of more than one year. . . .

CHAPTER III:

Section 591 (b) Federal Regulations

In all other matters relating to the details of planning, developing, implementing, and evaluating programs and projects by State and local educational agencies, the Secretary shall not issue regulations, but may consult with appropriate State, local, and private educational agencies and, upon request, provide technical assistance, information, and suggested guidelines designed to promote the development and implementation of effective instructional programs and to otherwise assist in carrying out the purposes of this subtitle.

1/ The Education Consolidation and Improvement Act of 1981 was enacted as part of P. L. 97-35.

APPENDIX D

Planning and Advisory Group for the Assessment of TIERS

Robert Stonehill, Chair
Department of Education
Office of Planning, Budget
and Evaluation
Washington, D.C.

Robert Cheuvront
Coordinator, Title I
Colorado Department of Education
Denver, Colorado

Gary Echternacht
Title I Technical Assistance Center
Region II
Educational Testing Service
Princeton, New Jersey

Joy Frechtling
Office of Program Assessment
Montgomery County Public Schools
Rockville, Maryland

Oliver Himley
Coordinator, Title I
Iowa Department of Public Instruction
Des Moines, Iowa

Thomas Krueck
Title I Technical Assistance Center
Region VI
Powell Associates
Austin, Texas

Claudia Merkel-Keller
Director of Evaluation
New Jersey Department of Education
Trenton, New Jersey

William Padia
Office of Planning, Evaluation
and Research
California Department of Education
Sacramento, California

Paul Rost
Title I Evaluator
Albuquerque Public Schools
Albuquerque, New Mexico

Ex Officio

Judith Anderson
Department of Education
Office of Planning, Budget
and Evaluation
Washington, D.C.

Thomas Enderlein
Department of Education
Office of Compensatory Education
Programs
Washington, D.C.

James English
Department of Education
Office of Planning, Budget
and Evaluation
Washington, D.C.

William Lobosco
Department of Education
Office of Compensatory Education
Programs
Washington, D.C.

APPENDIX E

SEAs and LEAs in Which Interviews Were Conducted for This Study

Assistance in conducting this study was provided by educational administrators and evaluators associated with the following SEAs:

Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
District of Columbia
Florida
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oregon
Pennsylvania
Rhode Island
South Carolina
Tennessee
Texas
Utah
Vermont
Washington
West Virginia
Wyoming

In order to preserve the anonymity of LEA respondents, the following list indicates states in which LEA interviews were conducted. At least one LEA was interviewed in each of these states.

Arizona
Arkansas
California
Colorado
Illinois
Iowa
Louisiana
Maine
Massachusetts
Michigan
Montana
Nebraska
New Hampshire
New Mexico
New York
Ohio
Pennsylvania
Rhode Island
South Carolina
Tennessee
Texas
Utah
Virginia

APPENDIX F

Federal Personnel Interviewed for This Study

A number of individuals in Washington were interviewed as part of this study. They included staff of the U.S. Department of Education and the U.S. Congress. We wish to thank all of them. They are listed below in alphabetical order, along with the office with which they are affiliated:

Albert Alford
ED Office of the Secretary
(formerly, Office of Legislation)

Janice Anderson
ED Office of Planning, Budget and Evaluation

Judith Anderson
ED Office of Planning, Budget and Evaluation

Lawrence Brown
ED Office of Planning, Budget and Evaluation

Sandy Brown
ED Office of Planning, Budget and Evaluation

Richard DiEugenio
House Education and Labor Committee

Thomas Enderlein
ED Office of Compensatory Education Programs

John Evans
Educational Testing Service in Berkeley, Cal.
(formerly OE Office of Planning, Budget and Evaluation)

Eugene Fusco
ED Office of Compensatory Education Programs

Edward Glassman
ED Office of Planning, Budget and Evaluation

Carolyn Horner
ED Office of Compensatory Education Programs

James Hyman
ED Office of the Deputy Undersecretary for Planning, Budget and Evaluation

John Jennings
House Education and Labor Committee

William Lobosco
ED Office of Compensatory Education Programs

Robert Maroney
ED Office of Planning, Budget and Evaluation

Helen O'Leary
ED Chapter II Task Force
(formerly Title II Basic Skills Office)

Ricky Poster
Senate Appropriations Committee

Richard Prager
House Budget Committee

Benjamin Rice
ED Office of Compensatory Education Programs

David Shoemaker
ED Office of Planning, Budget and Evaluation

Dorothy Shuler
ED Office of Planning, Budget and Evaluation

Robert Stonehill
ED Office of Planning, Budget, and Evaluation

Carl Wisler
U. S. Government Accounting Office
(formerly OE Office of Planning, Budget and Evaluation)